

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 14:58:10 ; Search time 2878.96 Seconds  
(without alignments)  
218.064 Million cell updates/sec

Title: US-09-623-329-5  
 Perfect score: 30  
 Sequence: 1 catctcccccttqqaatqagccctqagccc 30

Scoring table: IDENTITY\_NUC  
Gapop 10.0 : Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues  
Total number of hits satisfying chosen parameters: 3595312

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

Database : GenEmbl : \*

```

1:  gb:ba:*
2:  gb:htg:*
3:  gb:in:*
4:  gb:om:*
5:  gb:ov:*
6:  gb:pat:*
7:  gb:ph:*
8:  gb:pl:*
9:  gb:pr:*
10: gb:ro:*
11: gb:st:*
12: gb:sys:*
13: gb:un:*
14: gb:vi:*
15: em:ba:*
16: em:fun:*
17: em:hum:*
18: em:in:*
19: em:mu:*
20: em:om:*
21: em:or:*
22: em:ov:*
23: em:pat:*
24: em:ph:*
25: em:pl:*
26: em:ro:*
27: em:st:*
28: em:un:*
29: em:vi:*
30: em:htg:hum:*
31: em:htg:inv:*
32: em:htg:other:*
33: em:htg:inv:

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Query	Score	Length	ID	Description
No.	Match				

1	30	100.0	30	AX018446	Sequence
2	30	100.0	627	14	U021193
3	30	100.0	627	14	U021194
4	30	100.0	627	14	U021195
5	30	100.0	627	14	U021196
6	30	100.0	627	14	U021197
7	30	100.0	627	14	U021198
8	30	100.0	627	14	U021199
9	30	100.0	627	14	U021200
10	30	100.0	627	14	U021201
11	30	100.0	627	14	U021202
12	30	100.0	627	14	U021203
13	30	100.0	627	14	U021204
14	30	100.0	636	14	U021205
15	30	100.0	1748	6	AK116267
16	30	100.0	1926	6	AK107940
17	30	100.0	2580	6	AK108994
18	30	100.0	5452	6	AK081151
19	30	100.0	5452	12	U02454
20	30	100.0	9600	6	A92665
21	30	100.0	9600	6	AK158345
22	30	100.0	10596	6	125041
23	30	100.0	10596	6	140503
24	30	100.0	10737	12	xx002428
25	30	100.0	10850	12	U02455
26	30	100.0	11265	6	AX018159
27	30	100.0	172281	14	FBV
28	30	100.0	184113	14	HS418958KAJ
29	21	70.0	30387	2	AC078987
30	21	70.0	36865	9	HSV87088
31	21	70.0	40584	9	AL133348
32	21	70.0	41029	2	HSUR081
33	21	70.0	180129	2	AC094439
34	21	70.0	232784	2	AC106528
35	20.6	68.7	127384	9	AL157823
36	20.6	68.7	159482	2	AL512371
37	20.4	68.0	2613	9	AK022290
38	20.4	68.0	150113	9	AC015988
39	20.4	68.0	162274	2	AC013819
40	20.4	68.0	166075	2	AC092836
41	20.4	68.0	189008	2	AP001491
42	20.4	68.0	189241	2	AC090402
43	20.4	68.0	200820	2	AC107896
44	20.2	67.3	156856	2	AC018521
45	20.2	66.7	90767	2	H0510A06

## ALIGNMENTS

```

RESULT      1
AX018446   30 bp DNA
LOCUS      1 linear
DEFINITION Herpesvirus 4
ACCESSION  AF094515.5
VERSION    1
KEYWORDS   Epstein-Barr virus,
SOURCE     Human herpesvirus 4
ORGANISM   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
            Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE  1 (bases 1 to 30)
AUTHORS   Middlejorp,J.M., Van Den Brule,A.J. and Vervooft,M.B.
TITLE     Oligonucleotides for the amplification and detection of Epstein
          Barr virus (ebv) nucleic acid
JOURNAL    Patent: WO 94/5155-A 5 10-SEP-1999;
MIDDELDRUP JAAP MICHEL (NL); AKZO NOBEL NV (NL); DEN BRULE
ADRIANUS JOHANNES CH (NL); VERVOORT MARCEL BARTOLINA HEND (NL)
FEATURES   Location/Qualifiers
            1..30
                /organism="Human herpesvirus 4"
                /db_xref="taxon:10376"
BASE COUNT 3 a 13 c 7 g 1 t

```

```

ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.0033;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgtctcccttggatggccctgacc 30
|||||
Db 1 CGTCTCCCTTGGATGGCCCTGGACC 30

RESULT 2
HHU21193
LOCUS Human herpesvirus 4 627 bp DNA linear VRL 24-MAY-1995
DEFINITION Human herpesvirus 4, isolate AM nuclear antigen EBNA-1 gene,
partial cds, C-terminal unique domain.
ACCESSION U21193
VERSION U21193.1 GI:710371
SOURCE Epstein-Barr virus.
ORGANISM Human herpesvirus 4
VIRUSES: dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE 1 (bases 1 to 609)
AUTHORS Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,
Rooney,C.M. and Arrand,J.R.
TITLE Antigenic and sequence variation in the C-terminal unique domain of
the Epstein-Barr virus nuclear antigen EBNA-1
JOURNAL Virology (1995) In press
REFERENCE 2 (bases 1 to 627)
AUTHORS Pepper,S.D.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for
Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,
Manchester, Britain, M20 9HX
FEATURES
location/Qualifiers
1..627
/organism="Human herpesvirus 4"
/isolate="Ag876"
/specific_host="Homo sapiens"
/db_xref="taxon:10376"
<1..612
/note="C-terminal unique domain"
/codon_start=1
/product="nuclear antigen EBNA-1"
/protein_id="AAA67275.1"
/db_xref="GI:710374"
/translation="ADDPGEGPSTGPRGGGRRKKGGWFGKRGEGSSQKPFENIA
EGLRLLARSHVERTTDEGNVAGVFYGGSKTSLYNLRGIGLAI PQCRLTPLSRLP
FGMAPGPGQPLRESIVCYFIVLQTHIFAEGLKDAIKDLVLPKPAPTCNIKVTVC
SFDDGVDLPPWFPMPVMEGAAGAGDGGDGGDEGESEGE"
BASE COUNT 152 a 128 c 196 g 151 t
ORIGIN

Query Match 100.0%; Score 30; DB 14; Length 627;
Best Local Similarity 100.0%; Pred. No. 0.0033;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgtctcccttggatggccctgacc 30
|||||
Db 298 CGTCTCCCTTGGATGGCCCTGGACC 327

RESULT 4
HHU21195
LOCUS Human herpesvirus 4 627 bp DNA linear VRL 24-MAY-1995
DEFINITION Human herpesvirus 4, isolate EG nuclear antigen EBNA-1 gene,
partial cds, C-terminal unique domain.
ACCESSION U21195
VERSION U21195.1 GI:710375
KEYWORDS
SOURCE Epstein-Barr virus.
ORGANISM Human herpesvirus 4
VIRUSES: dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE 1 (bases 1 to 609)
AUTHORS Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,
Rooney,C.M. and Arrand,J.R.
TITLE Antigenic and sequence variation in the C-terminal unique domain of
the Epstein-Barr virus nuclear antigen EBNA-1
JOURNAL Virology (1995) In press
REFERENCE 2 (bases 1 to 627)
AUTHORS Pepper,S.D.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for
Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,
Manchester, Britain, M20 9BX
FEATURES
location/Qualifiers
1..627
/organism="Human herpesvirus 4"
/isolate="EG"
/specific_host="Homo sapiens"
/db_xref="taxon:10376"
<1..612
/note="C-terminal unique domain"
/codon_start=1
/product="nuclear antigen EBNA-1"
/protein_id="AAA67275.1"
/db_xref="GI:710372"
/translation="ADDPGEGPSTGPRGGGRRKKGGWFGKRGEGSSQKPFENIA
EGLRLLARSHVERTTDEGNVAGVFYGGSKTSLYNLRGIGLAI PQCRLTPLSRLP
FGMAPGPGQPLRESIVCYFIVLQTHIFAEGLKDAIKDLVLPKPAPTCNIKVTVC
SFDDGVDLPPWFPMPVMEGAAGAGDGGDGGDEGESEGE"
BASE COUNT 154 a 133 c 194 g 147 t
ORIGIN

Query Match 100.0%; Score 30; DB 14; Length 627;
Best Local Similarity 100.0%; Pred. No. 0.0033;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgtctcccttggatggccctgacc 30
|||||
Db 298 CGTCTCCCTTGGATGGCCCTGGACC 327

RESULT 3
HHU21194
LOCUS Human herpesvirus 4 627 bp DNA linear VRL 24-MAY-1995
DEFINITION Human herpesvirus 4, isolate Ag876 nuclear antigen EBNA-1 gene,
partial cds, C-terminal unique domain.
ACCESSION U21194
VERSION U21194.1 GI:710373
KEYWORDS

```

```

/specific_host-"Homo sapiens"
/db_xref="taxon:10376"
<!. .612
/notes-"C-terminal unique domain"
/codon_start=1
/product-"nuclear antigen EBNA-1"
/protein_id="AA67277.1"
/db_xref="GI:710376"
/translation-"ADDPGEGPSTGPRGOGDGRKKGGWFGKRRKGGSNPKFENIA
EGLRLALRCHVETEDGNWAGVYGGSKTSLYNLRGRTALAIPQCKRLTPLSLRP
FGMAPGPGPGLRESIVCYFMVFLQTHFAEVLKDAIKDLVMPKPAPTCNIRVTVC
SPDGDVLPWPFPMPWEGAAAGDDGDDGDEGGDEGEEOE"
BASE COUNT 154 a 132 c 195 g 146 t
ORIGIN
Query Match 100.0%; Score 40; DB 14; Length 627;
Best Local Similarity 100.0%; Pred. No. 0.0033;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 cgtctcccttgaatggccctgaacc 30
|||||
Db 298 CGTCTCCCTTTGGAATGGCCCTGGACCC 327

RESULT 6
HHU21197
LOCUS
DEFINITION
Human herpesvirus 4, isolate JF2 nuclear antigen EBNA-1 gene,
partial cds, C-terminal unique domain.
ACCESSION
U21197
VERSION
U21197.1 GI:710379
KEYWORDS
Epstein-Barr virus.
SOURCE
Human herpesvirus 4
ORGANISM
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
1 (bases 1 to 609)
Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,
Rooney,C.M. and Arrand,J.R.
Antigenic and sequence variation in the C-terminal unique domain of
the Epstein-Barr virus nuclear antigen EBNA-1
Virology (1995) In press
2 (bases 1 to 627)
Pepper,S.D.
Direct Submission
Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for
Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,
Manchester, Britain, M20 9BX
FEATURES
Location/Qualifiers
1..627
/oranism="Human herpesvirus 4"
/isolate="JF2"
/specific_host-"Homo sapiens"
/db_xref="taxon:10376"
<!. .612
/notes-"C-terminal unique domain"
/codon_start=1
/product-"nuclear antigen EBNA-1"
/protein_id="AA67279.1"
/db_xref="GI:710380"
/translation-"ADDPGEGPSTGPRGOGDGRKKGGWFGKRRKGGSNPKFENIA
EGLRLALRCHVETEDGNWAGVYGGSKTSLYNLRGRTALAIPQCKRLTPLSLRP
FGMAPGPGPGLRESIVCYFMVFLQTHFAEVLKDAIKDLVMPKPAPTCNIRVTVC
SPDGDVLPWPFPMPWEGAAAGDDGDDGDEGGDEGEEOE"
BASE COUNT 154 a 130 c 194 g 149 t
ORIGIN
Query Match 100.0%; Score 40; DB 14; Length 627;
Best Local Similarity 100.0%; Pred. No. 0.0033;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 cgtctcccttgaatggccctgaacc 30
|||||
Db 298 CGTCTCCCTTTGGAATGGCCCTGGACCC 327

RESULT 7
HHU21198
LOCUS
DEFINITION
Human herpesvirus 4, isolate LA nuclear antigen EBNA-1 gene,
partial cds, C-terminal unique domain.
ACCESSION
U21198
VERSION
U21198.1 GI:710381
KEYWORDS
Epstein-Barr virus.
SOURCE
Human herpesvirus 4
ORGANISM
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
1 (bases 1 to 609)
Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,
Rooney,C.M. and Arrand,J.R.
Antigenic and sequence variation in the C-terminal unique domain of
the Epstein-Barr virus nuclear antigen EBNA-1
Virology (1995) In press
2 (bases 1 to 627)
Pepper,S.D.
Direct Submission
Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for
Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,
Manchester, Britain, M20 9BX
FEATURES
Location/Qualifiers
1..627
/oranism="Human herpesvirus 4"
/isolate="JF1"
/specific_host-"Homo sapiens"
/db_xref="taxon:10376"
<!. .612
/notes-"C-terminal unique domain"
/codon_start=1
/product-"nuclear antigen EBNA-1"
/protein_id="AA67278.1"
/db_xref="GI:710378"
/translation-"ADDPGEGPSTGPRGOGDGRKKGGWFGKRRKGGSNPKFENIA
EGLRLALRCHVETEDGNWAGVYGGSKTSLYNLRGRTALAIPQCKRLTPLSLRP
FGMAPGPGPGLRESIVCYFMVFLQTHFAEVLKDAIKDLVMPKPAPTCNIRVTVC
SPDGDVLPWPFPMPWEGAAAGDDGDDGDEGGDEGEEOE"
BASE COUNT 151 a 128 c 197 g 151 t
ORIGIN
Query Match 100.0%; Score 30; DB 14; Length 627;
Best Local Similarity 100.0%; Pred. No. 0.0033;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 cgtctcccttgaatggccctgaacc 30
|||||
Db 298 CGTCTCCCTTTGGAATGGCCCTGGACCC 327

```

TITLE Antigenic and sequence variation in the C-terminal unique domain of the Epstein-Barr virus nuclear antigen EBNA-1  
JOURNAL Virology (1995) In press  
REFERENCE 2 (bases 1 to 627)  
AUTHORS Pepper, S.D.

TITLE Direct Submission  
JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd, Manchester, Britain, M20 9BX  
FEATURES Location/Qualifiers

source 1..627  
/organism="Human herpesvirus 4"  
/isolate="LA"  
/specific\_host="Homo sapiens"  
/db\_xref="taxon:10376"  
CDS <1..612  
/note="C-terminal unique domain"  
/codon\_start=1  
/product="nuclear antigen EBNA-1"  
/protein\_id="AA67280.1"  
/db\_xref="GI:710382"

/translation="ADDPGEVSTGPRGGDGRKKKGWFGKRGGSNOKFENIA  
EGLRTLLARCHVERTDEGTWAGVYGGSKTSLYNLKRGIALAIPQCRITPLSRILP  
FGMAPGPGPGLRESIVCYFIVFLQTHIFAEGLKDAIKDLVMPKPAPTCNIKATVC  
SFDDGVDLPPWFPMPWVEGAAAGDGDGDDGGGDEGEQGE"  
BASE COUNT 155 a 132 c 191 q 149 t

ORIGIN

Query Match 100.0%; Score 30; DB 14; Length 627;  
Best Local Similarity 100.0%; Pred. No. 0.0033;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgtctcccttggatggccctggacc 30  
|||||  
DB 298 CGTCTCCCTTGGATGGCCCTGGACC 327

RESULT 8

HHU21199  
LOCUS Human herpesvirus 4, isolate M7 nuclear antigen EBNA-1 gene,  
DEFINITION partial cds, C-terminal unique domain.  
ACCESSION U21199  
VERSION 2  
KEYWORDS 1 GI:710383  
SOURCE Epstein-Barr virus.  
ORGANISM Human herpesvirus 4  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Gammaherpesvirinae; Lymphocryptovirus.

REFERENCE 1 (bases 1 to 609)  
AUTHORS Wright, M.N., Stewart, J.P., Janjua, N.J., Pepper, S.D., Sample, C., Rooney, C.M. and Arrand, J.R.  
TITLE Antigenic and sequence variation in the C-terminal unique domain of the Epstein-Barr virus nuclear antigen EBNA-1  
JOURNAL Virology (1995) In press  
REFERENCE 2 (bases 1 to 627)  
AUTHORS Pepper, S.D.

TITLE Direct Submission  
JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd, Manchester, Britain, M20 9BX  
FEATURES Location/Qualifiers

source 1..627  
/organism="Human herpesvirus 4"  
/isolate="M7"  
/specific\_host="Homo sapiens"  
/db\_xref="taxon:10376"  
CDS <1..612  
/note="C-terminal unique domain"  
/codon\_start=1  
/product="nuclear antigen EBNA-1"  
/protein\_id="AA67281.1"

/db\_xref="GI:710384"  
/translation="ADDPGEVSTGPRGGDGRKKKGWFGKRGGSNOKFENIA  
EGLRTLLARCHVERTDEGTWAGVYGGSKTSLYNLKRGIALAIPQCRITPLSRILP  
FGMAPGPGPGLRESIVCYFIVFLQTHIFAEGLKDAIKDLVMPKPAPTCNIKATVC  
SFDDGVDLPPWFPMPWVEGAAAGDGDGDDGGGDEGEQGE"  
BASE COUNT 153 a 141 c 193 g 150 t

ORIGIN

Query Match 100.0%; Score 30; DB 14; Length 627;  
Best Local Similarity 100.0%; Pred. No. 0.0033;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgtctcccttggatggccctggacc 30  
|||||  
DB 298 CGTCTCCCTTGGATGGCCCTGGACC 327

RESULT 9

HHU21200  
LOCUS Human herpesvirus 4, isolate M1 nuclear antigen EBNA-1 gene,  
DEFINITION partial cds, C-terminal unique domain.  
ACCESSION U21200  
VERSION 1  
KEYWORDS 1 GI:710385  
SOURCE Epstein-Barr virus.  
ORGANISM Human herpesvirus 4  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Gammaherpesvirinae; Lymphocryptovirus.

REFERENCE 1 (bases 1 to 609)  
AUTHORS Wright, M.N., Stewart, J.P., Janjua, N.J., Pepper, S.D., Sample, C., Rooney, C.M. and Arrand, J.R.  
TITLE Antigenic and sequence variation in the C-terminal unique domain of the Epstein-Barr virus nuclear antigen EBNA-1  
JOURNAL Virology (1995) In press  
REFERENCE 2 (bases 1 to 627)  
AUTHORS Pepper, S.D.

TITLE Direct Submission  
JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd, Manchester, Britain, M20 9BX  
FEATURES Location/Qualifiers

source 1..627

/organism="Human herpesvirus 4"  
/isolate="M1"  
/specific\_host="Homo sapiens"  
/db\_xref="taxon:10376"  
CDS <1..612  
/note="C-terminal unique domain"

/codon\_start=1  
/product="nuclear antigen EBNA-1"  
/protein\_id="AA67282.1"  
/db\_xref="GI:710386"  
/translation="ADDPGEVSTGPRGGDGRKKKGWFGKRGGSNOKFENIA  
EGLRTLLARCHVERTDEGTWAGVYGGSKTSLYNLKRGIALAIPQCRITPLSRILP  
FGMAPGPGPGLRESIVCYFIVFLQTHIFAEGLKDAIKDLVLPKPAPTCNIKATVC  
SFDDGVDLPPWFPMPWVEGAAAGDGDGDDGGGDEGEQGE"  
BASE COUNT 154 a 133 c 192 g 148 t

Query Match 100.0%; Score 30; DB 14; Length 627;  
Best Local Similarity 100.0%; Pred. No. 0.0033;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgtctcccttggatggccctggacc 30  
|||||  
DB 298 CGTCTCCCTTGGATGGCCCTGGACC 327

RESULT 10

HHU21201



```

LOCUS           HHU21201               627 bp    DNA          linear      VRL 24 MAY-1995
DEFINITION      Human herpesvirus 4, isolate p3hr1 nuclear antigen EBNA-1 gene,
ACCESSION       U21201
VERSION         U21201.1  GI:710387
KEYWORDS
SOURCE
ORGANISM        Epstein-Barr virus.
                Human herpesvirus 4
                Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
                Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE
AUTHORS         Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,
                Rooney,C.M. and Arrand,J.R.
TITLE           Antigenic and sequence variation in the C-terminal unique domain of
                the Epstein-Barr virus nuclear antigen EBNA-1
JOURNAL         Virology (1995) In press
AUTHORS         2 (bases 1 to 627)
                Pepper,S.D.
DIRECT SUBMISSION
SUBMITTED (16-FEB-1995) Stuart D. Pepper, Paterson Institute for
Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,
Manchester, Britain, M20 9BX
FEATURES        Location/Qualifiers
                source
                1..627
                /organism="Human herpesvirus 4"
                /isolate="p3hr1"
                /specific_host="Homo sapiens"
                /db_xref="taxon:10476"
                <1..612
                /note="C-terminal unique domain"
                /codon_start=1
                /product="nuclear antigen EBNA-1"
                /protein_id="AAA67284.1"
                /db_xref="GI:710488"
                /translation="ADDPGEFSTGPRGQDGKGRKKKGWFGKRGCGGSGNOKFENIA
                EGRLLARSHVERITDEGTWAGVYVGGSKTSLYNLKRGIALAIPOCRLLPLSKLP
                FGMAPGPGPQLRESIVCYFIVFLQTHFAEGLKDAIKDLVLPKPAPTCNIRVTVC
                SFDDGVDLPPWFPFMVEGAAGDGDGDEGEDEGEQGE"
BASE COUNT      151 a 129 c 196 q 151 t
ORIGIN
Query Match    100.0%; Score 30; DB 14; Length 627;
Best Local Similarity 100.0%; Pred. No. 0.0033;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 cgtctcccttgaatgagccctgaacc 30
Db 298 CGTCCCGCTTGATGATGCGCCCTGGAGTC 427

LOCUS           HHU21202               627 bp    DNA          linear      VRL 24 MAY-1995
DEFINITION      Human herpesvirus 4, isolate PA nuclear antigen EBNA-1 gene,
ACCESSION       U21202
VERSION         U21202.1  GI:710389
KEYWORDS
SOURCE
ORGANISM        Epstein-Barr virus.
                Human herpesvirus 4
                Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
                Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE
AUTHORS         Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,
                Rooney,C.M. and Arrand,J.R.
TITLE           Antigenic and sequence variation in the C-terminal unique domain of
                the Epstein-Barr virus nuclear antigen EBNA-1
JOURNAL         Virology (1995) In press
AUTHORS         2 (bases 1 to 627)
                Pepper,S.D.
DIRECT SUBMISSION
SUBMITTED (16-FEB-1995) Stuart D. Pepper, Paterson Institute for

```

```

Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,
Manchester, Britain, M20 9BX
FEATURES        Location/Qualifiers
                source
                1..627
                /organism="Human herpesvirus 4"
                /isolate="PA"
                /specific_host="Homo sapiens"
                /db_xref="taxon:10376"
                <1..612
                /note="C-terminal unique domain"
                /codon_start=1
                /product="nuclear antigen EBNA-1"
                /protein_id="AAA67284.1"
                /db_xref="GI:710390"
                /translation="ADDPGEFSTGPRGQDGKGRKKKGWFGKRGCGGSGNOKFENIA
                EGRLLARSHVERITDEGTWAGVYVGGSKTSLYNLKRGIALAIPOCRLLPLSKLP
                FGMAPGPGPQLRESIVCYFIVFLQTHFAEGLKDAIKDLVLPKPAPTCNIRVTVC
                SFDDGVDLPPWFPFMVEGAAGDGDGDEGEDEGEQGE"
BASE COUNT      155 a 130 c 193 q 149 t
ORIGIN
Query Match    100.0%; Score 40; DB 14; Length 627;
Best Local Similarity 100.0%; Pred. No. 0.0033;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 cgtctcccttgaatgagccctgaacc 30
Db 298 CGTCCCGCTTGATGATGCGCCCTGGAGTC 427

LOCUS           HHU21203               627 bp    DNA          linear      VRL 24-MAY-1995
DEFINITION      Human herpesvirus 4, isolate PM nuclear antigen EBNA-1 gene,
ACCESSION       U21203
VERSION         U21203.1  GI:710391
KEYWORDS
SOURCE
ORGANISM        Epstein-Barr virus.
                Human herpesvirus 4
                Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
                Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE
AUTHORS         Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,
                Rooney,C.M. and Arrand,J.R.
TITLE           Antigenic and sequence variation in the C-terminal unique domain of
                the Epstein-Barr virus nuclear antigen EBNA-1
JOURNAL         Virology (1995) In press
AUTHORS         2 (bases 1 to 627)
                Pepper,S.D.
DIRECT SUBMISSION
SUBMITTED (16-FEB-1995) Stuart D. Pepper, Paterson Institute for
Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,
Manchester, Britain, M20 9BX
FEATURES        Location/Qualifiers
                source
                1..627
                /organism="Human herpesvirus 4"
                /isolate="PM"
                /specific_host="Homo sapiens"
                /db_xref="taxon:10376"
                <1..612
                /note="C-terminal unique domain"
                /codon_start=1
                /product="nuclear antigen EBNA-1"
                /protein_id="AAA67285.1"
                /db_xref="GI:710392"
                /translation="ADDPGEFSTGPRGQDGKGRKKKGWFGKRGCGGSGNOKFENIA
                EGRLLARSHVERITDEGTWAGVYVGGSKTSLYNLKRGIALAIPOCRLLPLSKLP
                FGMAPGPGPQLRESIVCYFIVFLQTHFAEGLKDAIKDLVLPKPAPTCNIRATVC
                SFDDGVDLPPWFPFMVEGAAGDGDGDEGEDEGEQGE"
BASE COUNT      155 a 132 c 192 g 148 t
ORIGIN

```

Query Match 100.0%; Score 30; DB 14; Length 627;  
Best Local Similarity 100.0%; Pred. No. 0.0033;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 cgtctcccttgggaatgcccctggacc 30  
|||||  
Db 298 CGTCTCCCTTTGGGAATGCCCTGGACCC 327

RESULT 13  
HHU21204  
LOCUS Human herpesvirus 4, isolate SB nuclear antigen EBNA-1 gene, VRL 24-MAY-1995  
DEFINITION partial cds, C-terminal unique domain.  
ACCESSION U21204  
VERSION U21204.1 GI:710393  
KEYWORDS  
SOURCE  
ORGANISM  
Human herpesvirus 4  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Gammaherpesvirinae; Lymphocryptovirus.  
REFERENCE  
AUTHORS Wright, M.N., Stewart, J.P., Janjua, N.J., Pepper, S.D., Sample, C., Rooney, C.M. and Arrand, J.R.  
TITLE Antigenic and sequence variation in the C-terminal unique domain of the Epstein-Barr virus nuclear antigen EBNA-1  
JOURNAL Virology (1995) In press  
REFERENCE  
AUTHORS Pepper, S.D.  
TITLE Direct Submission  
JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd, Manchester, Britain, M20 9BX  
Location/Qualifiers  
1. 627  
/organism="Human herpesvirus 4"  
/isolate="SB"  
/specific\_host="Homo sapiens"  
/db\_xref="taxon:10476"  
<1. 612  
/note="C-terminal unique domain"  
/codon\_start=1  
/product="nuclear antigen EBNA-1"  
/protein\_id="AAA67286.1"  
/db\_xref="GI:710394"  
/translation="ADDPGEGPSTGPRGQGGRRKKGWFGKRGCGGSGNPKFENIA  
DGLRALLARSHVERTDECTWAGVYVGGSKTSYLNRRGTLALIPQCLPLSLRLP  
FGMAPGPGQPLRESIVCYFVPLQTHIFAEGLKDALIKDLVMPKPAFTCNIRVTC  
SFDDGVDLPPWFPFPMVGAAGAGDGDGDDGDEGDEGEQE"

BASE COUNT 153 a 132 c 194 g 148 t  
ORIGIN  
Query Match 100.0%; Score 30; DB 14; Length 627;  
Best Local Similarity 100.0%; Pred. No. 0.0033;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 cgtctcccttgggaatgcccctggacc 30  
|||||  
Db 298 CGTCTCCCTTTGGGAATGCCCTGGACCC 327

RESULT 14  
HHU21205  
LOCUS Human herpesvirus 4, isolate WW nuclear antigen EBNA-1 gene, VRL 24-MAY-1995  
DEFINITION partial cds, C-terminal unique domain.  
ACCESSION U21205  
VERSION U21205.1 GI:710395  
KEYWORDS  
SOURCE  
ORGANISM  
Epstein-Barr virus.  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Gammaherpesvirinae; Lymphocryptovirus.  
REFERENCE  
AUTHORS Wright, M.N., Stewart, J.P., Janjua, N.J., Pepper, S.D., Sample, C., Rooney, C.M. and Arrand, J.R.  
TITLE Antigenic and sequence variation in the C-terminal unique domain of the Epstein-Barr virus nuclear antigen EBNA-1  
JOURNAL Virology (1995) In press  
REFERENCE  
AUTHORS Pepper, S.D.  
TITLE Direct Submission  
JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd, Manchester, Britain, M20 9BX  
Location/Qualifiers  
1. 627  
/organism="Human herpesvirus 4"  
/isolate="SB"  
/specific\_host="Homo sapiens"  
/db\_xref="taxon:10476"  
<1. 612  
/note="C-terminal unique domain"  
/codon\_start=1  
/product="nuclear antigen EBNA-1"  
/protein\_id="AAA67286.1"  
/db\_xref="GI:710394"  
/translation="ADDPGEGPSTGPRGQGGRRKKGWFGKRGCGGSGNPKFENIA  
DGLRALLARSHVERTDECTWAGVYVGGSKTSYLNRRGTLALIPQCLPLSLRLP  
FGMAPGPGQPLRESIVCYFVPLQTHIFAEGLKDALIKDLVMPKPAFTCNIRVTC  
SFDDGVDLPPWFPFPMVGAAGAGDGDGDDGDEGDEGEQE"

BASE COUNT 153 a 132 c 194 g 148 t  
ORIGIN  
Query Match 100.0%; Score 30; DB 14; Length 627;  
Best Local Similarity 100.0%; Pred. No. 0.0033;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 cgtctcccttgggaatgcccctggacc 30  
|||||  
Db 298 CGTCTCCCTTTGGGAATGCCCTGGACCC 327

RESULT 14  
HHU21205  
LOCUS Human herpesvirus 4, isolate WW nuclear antigen EBNA-1 gene, VRL 24-MAY-1995  
DEFINITION partial cds, C-terminal unique domain.  
ACCESSION U21205  
VERSION U21205.1 GI:710395  
KEYWORDS  
SOURCE  
ORGANISM  
Epstein-Barr virus.  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Gammaherpesvirinae; Lymphocryptovirus.  
REFERENCE  
AUTHORS Wright, M.N., Stewart, J.P., Janjua, N.J., Pepper, S.D., Sample, C., Rooney, C.M. and Arrand, J.R.  
TITLE Antigenic and sequence variation in the C-terminal unique domain of the Epstein-Barr virus nuclear antigen EBNA-1  
JOURNAL Virology (1995) In press  
REFERENCE  
AUTHORS Pepper, S.D.  
TITLE Direct Submission  
JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd, Manchester, Britain, M20 9BX  
Location/Qualifiers  
1. 636  
/organism="Human herpesvirus 4"  
/isolate="WW"  
/specific\_host="Homo sapiens"  
/db\_xref="taxon:10476"  
<1. 621  
/note="C-terminal unique domain"  
/codon\_start=1  
/product="nuclear antigen EBNA-1"  
/protein\_id="AAA67287.1"  
/db\_xref="GI:710396"  
/translation="ADDPGEGPSTGPRGQGGRRKKGWFGKRGCGGSGNPKFENIA  
EGIRTLARSHVERTDECTWAGVYVGGSKTSYLNRRGTLALIPQCLPLSLRLP  
FGMAPGPGQPLRESIVCYFVPLQTHIFAEGLKDALIKDLVMPKPAFTCNIRVTC  
SFDDGVDLPPWFPFPMVGAAGAGDGDGDDGDEGDEGEQE"

ORGANISM Human herpesvirus 4  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Gammaherpesvirinae; Lymphocryptovirus.  
REFERENCE  
AUTHORS Wright, M.N., Stewart, J.P., Janjua, N.J., Pepper, S.D., Sample, C., Rooney, C.M. and Arrand, J.R.  
TITLE Antigenic and sequence variation in the C-terminal unique domain of the Epstein-Barr virus nuclear antigen EBNA-1  
JOURNAL Virology (1995) In press  
REFERENCE  
AUTHORS Pepper, S.D.  
TITLE Direct Submission  
JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd, Manchester, Britain, M20 9BX  
Location/Qualifiers  
1. 636  
/organism="Human herpesvirus 4"  
/isolate="WW"  
/specific\_host="Homo sapiens"  
/db\_xref="taxon:10476"  
<1. 621  
/note="C-terminal unique domain"  
/codon\_start=1  
/product="nuclear antigen EBNA-1"  
/protein\_id="AAA67287.1"  
/db\_xref="GI:710396"  
/translation="ADDPGEGPSTGPRGQGGRRKKGWFGKRGCGGSGNPKFENIA  
EGIRTLARSHVERTDECTWAGVYVGGSKTSYLNRRGTLALIPQCLPLSLRLP  
FGMAPGPGQPLRESIVCYFVPLQTHIFAEGLKDALIKDLVMPKPAFTCNIRVTC  
SFDDGVDLPPWFPFPMVGAAGAGDGDGDDGDEGDEGEQE"

BASE COUNT 158 a 134 c 196 g 148 t  
ORIGIN  
Query Match 100.0%; Score 30; DB 14; Length 636;  
Best Local Similarity 100.0%; Pred. No. 0.0033;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 cgtctcccttgggaatgcccctggacc 40  
|||||  
Db 298 CGTCTCCCTTTGGGAATGCCCTGGACCC 127

RESULT 15  
AR116267  
LOCUS Sequence 1 from patent US 6133025.  
DEFINITION AK116267  
ACCESSION AK116267  
VERSION AK116267.1 GI:14096589  
KEYWORDS  
SOURCE  
ORGANISM  
Unclassified.  
REFERENCE  
AUTHORS Seed, B.  
TITLE Compact  
JOURNAL Patent: US 6133025-A 1 17-OCT-2000;  
Location/Qualifiers  
1. 1748  
/organism="unknown"  
BASE COUNT 453 a 382 c 546 g 367 t  
ORIGIN

Query Match 100.0%; Score 30; DB 6; Length 1748;  
Best Local Similarity 100.0%; Pred. No. 0.0033;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 cgtctcccttgggaatgcccctggacc 30  
|||||  
Db 1407 CGTCTCCCTTTGGGAATGCCCTGGACCC 1436

RESULT 15  
AR116267  
LOCUS Sequence 1 from patent US 6133025.  
DEFINITION AK116267  
ACCESSION AK116267  
VERSION AK116267.1 GI:14096589  
KEYWORDS  
SOURCE  
ORGANISM  
Unclassified.  
REFERENCE  
AUTHORS Seed, B.  
TITLE Compact  
JOURNAL Patent: US 6133025-A 1 17-OCT-2000;  
Location/Qualifiers  
1. 1748  
/organism="unknown"  
BASE COUNT 453 a 382 c 546 g 367 t  
ORIGIN

Query Match 100.0%; Score 30; DB 6; Length 1748;  
Best Local Similarity 100.0%; Pred. No. 0.0033;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 cgtctcccttgggaatgcccctggacc 30  
|||||  
Db 1407 CGTCTCCCTTTGGGAATGCCCTGGACCC 1436

RESULT 15  
AR116267  
LOCUS Sequence 1 from patent US 6133025.  
DEFINITION AK116267  
ACCESSION AK116267  
VERSION AK116267.1 GI:14096589  
KEYWORDS  
SOURCE  
ORGANISM  
Unclassified.  
REFERENCE  
AUTHORS Seed, B.  
TITLE Compact  
JOURNAL Patent: US 6133025-A 1 17-OCT-2000;  
Location/Qualifiers  
1. 1748  
/organism="unknown"  
BASE COUNT 453 a 382 c 546 g 367 t  
ORIGIN

Search completed: May 18, 2002, 14:58:11  
Job time: 8188 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:22:09 ; Search time 8624.33 Seconds  
(without alignments)  
46.950 Million cell updates/sec

Title: US-09-623-329-5  
Perfect score: 30  
Sequence: 1 cgtcccttttggaatgagccctggacc 30

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	100.0	657	12 AG045201	AG045201 Pan trogl
2	30	100.0	688	12 AG126122	AG126122 Pan trogl
3	21.2	70.7	870	10 BF977466	BF977466 602147210
4	21	70.0	662	12 AQ489303	AQ489303 RPCI-11-2
5	20.8	69.3	989	10 BE797449	BE797449 601587766
6	20.6	68.7	279	9 BB092712	BB092712 BB092712
7	20.6	68.7	290	9 BB365071	BB365071 BB365071
8	20.6	68.7	418	9 AA076453	AA076453 zm91h03.r
9	20.6	68.7	572	9 AA543953	AA543953 vj69c02.r
10	20.4	68.0	324	10 C05316	C05316 C05316 Huma
11	20.4	68.0	437	10 C04457	C04457 C04457 Huma
12	20.4	68.0	471	12 AZ007999	AZ007999 RPCI-23-3
13	20.4	68.0	477	10 C04066	C04066 C04066 Huma
14	20.4	68.0	551	12 AQ389422	AQ389422 RPCI11-15
15	20.4	68.0	663	9 BB645391	BB645391 BB645391
16	20.4	68.0	681	10 BF790590	BF790590 602250291
17	20.4	68.0	778	10 BF575127	BF575127 602133411

18	20.4	68.0	856	10 BF694716	BF694716 602080787
19	20.4	68.0	914	10 BF671649	BF671649 602151505
20	20.4	68.0	979	10 BF569138	BF569138 602184549
21	20.4	68.0	1582	10 BG334257	BG334257 602461758
22	20.2	67.3	327	9 BB847651	BB847651 BB847651
23	20	66.7	505	9 BE192806	BE192806 EST331535
24	20	66.7	571	9 A1001088	A1001088 EST331535
25	20	66.7	682	10 BE368467	BE368467 601220551
26	20	66.7	778	10 BE194859	BE194859 603346033
27	20	66.7	863	10 BG922376	BG922376 602820294
28	20	66.7	935	10 BG872263	BG872263 602790877
29	19.6	65.3	270	9 BS595166	BS595166 BS595166
30	19.6	65.3	739	10 B1826163	B1826163 603076327
31	19.6	65.3	904	10 BF784231	BF784231 602108095
32	19.6	65.3	915	10 BF672720	BF672720 602152029
33	19.6	65.3	1055	10 B1552937	B1552937 603197864
34	19.4	64.7	255	10 H42182	H42182 y011f02.r1
35	19.4	64.7	266	9 AA514199	AA514199 nf60e09.s
36	19.4	64.7	279	9 BB176883	BB176883 BB176883
37	19.4	64.7	331	9 AA647757	AA647757 vp04b02.r
38	19.4	64.7	451	9 A1325466	A1325466 mp51a01.x
39	19.4	64.7	488	9 AA472298	AA472298 v01c08.r
40	19.4	64.7	518	9 A1181794	A1181794 ub87e01.r
41	19.4	64.7	549	12 AZ329273	AZ329273 1M0053M22
42	19.4	64.7	611	12 AZ015339	AZ015339 RPCI-23-3
43	19.4	64.7	848	10 BG260054	BG260054 602371639
44	19.4	64.7	899	10 BE731956	BE731956 601568972
45	19.4	64.7	904	12 CNS03RXW	AL257837 Tetraodon

## ALIGNMENTS

RESULT 1  
LOCUS AG045201 657 bp DNA linear GSS 02-NOV-2001  
DEFINITION Pan troglodytes DNA, clone: PTB-024A19.F, genomic survey sequence.  
ACCESSION AG045201  
VERSION AG045201.1 GI:16582018  
KEYWORDS GSS; GSS (genome survey sequence).  
SOURCE Pan troglodytes male lymphoblast DNA, clone\_lib:PTB Chimpanzee Male  
BAC Library clone:PTB-024A19.F.  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
1 (sites)  
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Totoki,Y., Watanabe,H. and Sakaki,Y.  
BAC end sequences of Library PTB  
Unpublished  
2 (bases 1 to 657)  
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Totoki,Y., Watanabe,H. and Sakaki,Y.  
Direct Submission  
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-2 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail:chimbpgsc@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,  
Tel:81-45-503-9111, Fax:81-45-503-9170)  
Clones are derived from the chimpanzee BAC library PTB This BAC end  
was generated during the R&D process and may have higher chance of  
clone tracking errors.  
PRIMERS  
Sequencing: -21M13  
LIBRARY  
Vector : pKSL45  
R.Site 1 : SacI  
R.Site 2 : SacI  
Location/Qualifiers  
1. .657  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="PTB-024A19.F"







SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT

human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 418)  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins  
M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Mardis, E., Moore  
B., Morris, M., Parsons, J., Prange, C., Rifkin, B., Rohlfing, T.,  
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisan, E.,  
Underwood, K., Wohldmann, P., Waterston, R., Wilson, R. and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
9704478  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
WARNING: There is evidence that suggests that the 384-well parent  
plate of this clone contains both human and mouse derived clones.  
Thus, the origin of this clone is uncertain. This caution should be  
kept in mind should you use this clone.  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 316.  
Location/Qualifiers  
FEATURES  
source  
1..418  
/organism="Homo sapiens"  
/db\_xref="GDB:3923521"  
/db\_xref="taxon:9606"  
/clone="IMAGE:545343"  
/clone\_lib="Stratagene ovarian cancer (#937219)"  
/sex="female"  
/dev\_stage="adult, 64 years"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: Bluescript SK-; Site\_1: EcoRI; Site\_2: XhoI  
; Cloned unidirectionally. Primer: Oligo dT. Papillary  
serous carcinoma, isolated from ascites, 64 year old  
caucasian. Average insert size: 0.8 kb; Uni-ZAP XR vector;  
-5' adaptor sequence: 5' GAATTCGACGACG 3' -3' adaptor  
sequence: 5' CTCGAGTTTTTTTTTTT 3'."  
BASE COUNT 76 a 127 c 144 g 60 t 11 others  
ORIGIN

Query Match 68.7%; Score 20.6; DB 9; Length 418;  
Best Local Similarity 79.3%; Pred. No. 2e+02; Indels 0; Gaps 0;  
Matches 23; Conservative 0; Mismatches 6;

Qy 2 gctcccttgggaatggcccttgacc 30  
||||| || ||||| || ||||| ||  
Db 266 GTCGCGCCCTTGAAGCGCTCTGGAGCC 238

RESULT 9  
AA543953  
LOCUS  
DEFINITION  
AA543953  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

572 bp mRNA linear EST 01-AUG-1997  
vJ69c02.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA  
clone IMAGE:934274 5' similar to gb:M22432 Mus musculus protein  
synthesis elongation factor Tu (MOUSE); mRNA sequence.  
AA543953.1 GI:2292430  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 572)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

TITLE  
JOURNAL  
COMMENT

Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MCL:539194  
High quality sequence stop: 337.  
Location/Qualifiers  
FEATURES  
source  
1..572  
/organism="Mus musculus"  
/strain="B6H2 F1/J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:934274"  
/clone\_lib="Knowles Solter mouse blastocyst B1"  
/tissue\_type="blastocyst"  
/dev\_stage="embryo (pre-implantation)"  
/lab\_host="bhl10B"  
/note="Organ: embryo; Vector: pSPORT; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally from mRNA prepared  
from 800 blastocysts. Primer: SalI(dT):  
5'-CGCTCGACCGTCGAC-3' TTTT-3'. cDNAs were  
cloned into the Nc I/SalI sites of a pSPORT vector (Life  
Technologies). Two different size selections: B1 (larger  
inserts) and B3."  
BASE COUNT 142 a 143 c 149 g 138 t  
ORIGIN

Query Match 68.7%; Score 20.6; DB 9; Length 572;  
Best Local Similarity 85.2%; Pred. No. 2.1e+02; Indels 0; Gaps 0;  
Matches 23; Conservative 0; Mismatches 4;

Qy 4 ctcccttgggaatggcccttgacc 30  
||||| || ||||| ||||| ||  
Db 518 CTCCCATCTGTAATGGCCCTGGAACC 544

RESULT 10  
C05316  
LOCUS  
DEFINITION  
C05316  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT

324 bp mRNA linear EST 30-JUL-1996  
C05316 Human heart cDNA (YNakamura) Homo sapiens cDNA clone  
3MHC011, mRNA sequence.  
C05316  
C05316.1 GI:1468567  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 324)  
Tanaka, T., Ogiwara, A., Uchiyama, I., Takagi, T., Yazaki, Y. and  
Nakamura, Y.  
Construction of a normalized directionally cloned cDNA library from  
adult heart and analysis of 3040 clones by partial sequencing  
Genomics 35 (1), 231-235 (1996)  
96299762  
Contact: Yusuke Nakamura  
Institute of Medical Science  
University of Tokyo  
4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan  
Tel: 81-3-5449-5372  
Fax: 81-3-5449-5433  
Email: yusuke@ims.u-tokyo.ac.jp.



```

FEATURES
  source
    Location/Qualifiers
      1..324
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone_lib="Human heart cDNA (YNakamura)"
        /dev_stage="adult"
        /note="Organ: heart; normalized directionally cloned cDNA
        from adult heart"
BASE COUNT      56 a      42 g      122 t      6 others
ORIGIN
Query Match      68.0%; Score 20.4; DB 10; Length 324;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 cgtctcccttggaatggccctggaacc 30
    ||| ||||| ||||| ||| ||||| ||
DB 191 CTCTCCCTTTGGAATGACTCTGGATCC 220

RESULT 11
LOCUS C04457
DEFINITION C04457 Human heart cDNA (YNakamura) Homo sapiens cDNA clone
ACCESSION 3NHC3361, mRNA sequence.
VERSION C04457
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 437)
Tanaka.T., Ogiwara.A., Uchiyama.I., Takagi.T., Yazaki.Y. and
Nakamura.Y.
TITLE Construction of a normalized directionally cloned cDNA library from
adult heart and analysis of 3040 clones by partial sequencing
JOURNAL Genomics 35 (1), 231-235 (1996)
MEDLINE 96299762
COMMENT Contact: Yusuke Nakamura
Institute of Medical Science
University of Tokyo
4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan
Tel: 81-3-5449-5372
Fax: 81-3-5449-5433
Email: yusuke@ims.u-tokyo.ac.jp.
FEATURES
  source
    Location/Qualifiers
      1..437
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone_lib="Human heart cDNA (YNakamura)"
        /dev_stage="adult"
        /note="Organ: heart; normalized directionally cloned cDNA
        from adult heart"
BASE COUNT      82 a      119 c      61 g      175 t
ORIGIN
Query Match      68.0%; Score 20.4; DB 10; Length 437;
Best Local Similarity 80.0%; Pred. No. 2.4e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 cgtctcccttggaatggccctggaacc 30
    ||| ||||| ||||| ||| ||||| ||
DB 194 CTCTCCCTTTGGAATGACTCTGGATCC 223

RESULT 12
LOCUS AZ007999
DEFINITION AZ007999 Mus musculus genomic clone RPCI-23-389123
ACCESSION RPCI-23-389123.TJ RPCI-23 Mus musculus genomic clone RPCI-23-389123
VERSION AZ007999.1 GI:7083383
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 471)
Zhao.S., Nierman.W., Feldblyum.T., Malek.J., Shatsman.S., Akinret
B., Levins.M., McGann.S., Tsegaye.G., Geer.K., Krol.M., de Jong.P.
and Fraser.C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
plate: 389 row: 1 column: 23
Seq primer: SP6
Class: BAC ends.
FEATURES
  source
    Location/Qualifiers
      1..471
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone_lib="RPCI-23-389123"
        /sex="female"
        /lab_host="DH10B"
        /note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
        EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
        brain genomic DNA was isolated and partially digested
        with a combination of EcoRI and EcoRI Methyase. Size
        selected DNA was cloned into the pBACE3.6 vector at the
        EcoRI sites. The ligation products were transformed into
        DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      132 a      94 c      156 g      87 t      2 others
ORIGIN
Query Match      68.0%; Score 20.4; DB 12; Length 471;
Best Local Similarity 95.5%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 5 tcccccttggaatggccctg 26
    ||||| ||||| ||||| |||||
DB 50 TCCCCTTTGGCATGGCCCTGG 71

RESULT 13
LOCUS C04066
DEFINITION C04066 Human heart cDNA (YNakamura) Homo sapiens cDNA clone
ACCESSION 3NHC2641, mRNA sequence.
VERSION C04066
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 477)

```

**AUTHORS** Tanaka,T., Ogiwara,A., Uchiyama,I., Takagi,T., Yazaki,Y. and Nakamura,Y.  
**TITLE** Construction of a normalized directionally cloned cDNA library from adult heart and analysis of 3040 clones by partial sequencing  
**JOURNAL** Genomics 35 (1), 231-235 (1996)  
**MEDLINE** 96299762  
**COMMENT** Contact: Yusuke Nakamura  
 Institute of Medical Science  
 University of Tokyo  
 4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan  
 Tel: 81-3-5449-5372  
 Fax: 81-3-5449-5433  
 Email: yusuke@ims.u-tokyo.ac.jp.

**FEATURES** source  
 1..477  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="JNHC2641"  
 /clone\_lib="Human heart cDNA (Y Nakamura)"  
 /dev\_stage="adult"  
 /note="Organ: heart; normalized directionally cloned cDNA from adult heart"

**BASE COUNT** 91 a 135 c 62 g 189 t  
**ORIGIN**

Query Match 68.0%; Score 20.4; DB 10; Length 477;  
 Best Local Similarity 80.0%; Pred. No. 2.5e+02;  
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 cgtctcccttgggaatggccctggaccc 30  
 I IIIIIIIIIIIIIIIIIIIIIIIIIIIII  
 Db 195 CTCTCTCCCTTGGGAATGACTCTGGATCC 224

**RESULT** 14  
**AQ389422/c**  
**LOCUS** AQ389422.1  
**DEFINITION** RPCI11-153F20-TV RPCI-11 Homo sapiens genomic clone RPCI-11-153F20, DNA sequence.  
**ACCESSION** AQ389422  
**VERSION** AQ389422.1 GI:4360445  
**KEYWORDS** GSS.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE** Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter J.C.  
**AUTHORS** Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter J.C.  
**TITLE** Use of BAC End Sequences from Library kPCI-11 for Sequence-Ready Map Building  
**JOURNAL** Unpublished (1997)  
**COMMENT** Contact: Shaying Zhao, William Nierman, Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: hbe@tigr.org  
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@jgmed.bufo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.bufo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html  
 Seq primer: T7  
 Class: BAC ends.

**FEATURES** source  
 1..551  
 /organism="Homo sapiens"  
 /db\_xref="GDB:7558507"  
 /db\_xref="taxon:9606"

/clone="RPCI-11-153F20"  
 /clone\_lib="RPCI-11"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; RPCI11 Human Male BAC Library"

**BASE COUNT** 175 a 116 c 121 g 139 t  
**ORIGIN**

Query Match 68.0%; Score 20.4; DB 12; Length 551;  
 Best Local Similarity 80.0%; Pred. No. 2.5e+02;  
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 cgtctcccttgggaatggccctggaccc 30  
 I IIIIIIIIIIIIIIIIIIIIIIIIIIIII  
 Db 275 CTCTCTCCCTTGGGAATGCGCTTGCAGCC 246

**RESULT** 15  
**BB645391**  
**LOCUS** BB645391  
**DEFINITION** Mus musculus cDNA clone B430115G18 5', mRNA sequence.  
**ACCESSION** BB645391  
**VERSION** BB645391.1 GI:16479858  
**KEYWORDS** EST.  
**SOURCE** house mouse.  
**ORGANISM** Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

**REFERENCE** 1 (bases 1 to 663)  
**AUTHORS** Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.  
**TITLE** RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
**JOURNAL** Unpublished (2001)  
**COMMENT** Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC) Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsr.riken.go.jp/  
 URL: http://genome.gsc.riken.go.jp/

**REFERENCE** Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
**AUTHORS** Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
**TITLE** Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
**JOURNAL** wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
**COMMENT** RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara Y. and Hayashizaki,Y.  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.  
 Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
 Please visit our web site (http://genome.gsc.riken.go.jp) for



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:22:02 ; Search time 8624.33 Seconds  
(without alignments)  
39.125 Million cell updates/sec

Title: US-09-623-329-3

Perfect score: 25

Sequence: 1 agagacaaggcttaatcgcatcc 25

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estnu:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_esti:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
C 1	21	84.0	657	12 AG045201	AG045201 Pan trogl
C 2	21	84.0	688	12 AG126122	AG126122 Pan trogl
C 3	19.2	76.8	255	12 BH222496	BH222496 1006107E0
C 4	19.2	76.8	412	9 AA190162	AA190162 mt95f04.r
C 5	19.2	76.8	459	9 AI450270	AI450270 mt95f04.x
C 6	19.2	76.8	507	9 AI509341	AI509341 mt95f04.y
C 7	18.8	75.2	593	10 BJ187698	BJ187698 BJ187698
C 8	18.8	75.2	631	10 BJ170162	BJ170162 BJ170162
C 9	18.8	75.2	674	10 BI894401	BI894401 ga54c11.y
C 10	18.6	74.4	365	10 BI186945	BI186945 AR046G08L
C 11	18.6	74.4	398	10 BI336254	BI336254 AR093A04L
C 12	18.6	74.4	601	12 AZ407701	AZ407701 1M0178J03
C 13	18.6	74.4	704	10 BF430776	BF430776 CG04E09P3
C 14	18.6	74.4	773	10 BI464363	BI464363 603204015
C 15	18.2	72.8	218	10 W91495	W91495 mg17e10.r1
C 16	18.2	72.8	253	10 BG650695	BG650695 sad98c12.
C 17	18.2	72.8	286	9 BB054250	BB054250 BB054250

C 18	18.2	72.8	300	9 BB005161	BB005161 BB005161
C 19	18.2	72.8	394	9 BB789045	BB789045 BB789045
C 20	18.2	72.8	450	9 AB832083	AB832083 sm19f06.y
C 21	18.2	72.8	456	10 BG651243	BG651243 sad99b01.
C 22	18.2	72.8	473	9 BE022084	BE022084 sm68c09.y
C 23	18.2	72.8	493	10 BM257857	BM257857 521722 MA
C 24	18.2	72.8	502	9 AA679566	AA679566 zj49p08.s
C 25	18.2	72.8	505	12 AZ863314	AZ863314 ZM0171G15
C 26	18.2	72.8	531	9 AW620941	AW620941 sj95g04.y
C 27	18.2	72.8	532	10 BG046711	BG046711 saa59b02.
C 28	18.2	72.8	547	10 BF750952	BF750952 202630 MA
C 29	18.2	72.8	588	9 AV174549	AV174549 AV174549
C 30	18.2	72.8	670	10 BI969664	BI969664 GM830008B
C 31	18.2	72.8	688	12 BH274915	BH274915 CH230-96H
C 32	18.2	72.8	706	10 BE533740	BE533740 601232888
C 33	18.2	72.8	955	10 BF690473	BF690473 602186938
C 34	17.8	71.2	187	10 BM108518	BM108518 EST556054
C 35	17.8	71.2	363	10 BI001050	BI001050 PM1-HN007
C 36	17.8	71.2	373	10 BG977318	BG977318 CM3-CS000
C 37	17.8	71.2	414	12 BH256030	BH256030 RDH10BAHO
C 38	17.8	71.2	415	10 BF153619	BF153619 031E03 Ma
C 39	17.8	71.2	448	10 BE924245	BE924245 EST428014
C 40	17.8	71.2	465	9 AA208439	AA208439 mv85b03.r
C 41	17.8	71.2	472	9 AA272220	AA272220 vb80a04.r
C 42	17.8	71.2	555	10 BE471411	BE471411 EST416264
C 43	17.8	71.2	565	10 BI436297	BI436297 EST5339058
C 44	17.8	71.2	633	10 BF459666	BF459666 063A06 Ma
C 45	17.8	71.2	637	10 BF459670	BF459670 063A12 Ma

## ALIGNMENTS

## RESULT 1

AG045201/c 657 bp DNA linear GSS 02-NOV-2001  
Pan troglodytes DNA, clone: PTB-024A19.F, genomic survey sequence.  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

AG045201  
Pan troglodytes DNA, clone: PTB-024A19.F, genomic survey sequence.  
AG045201  
GSS: GSS (genome survey sequence).  
Pan troglodytes male lymphoblast DNA, clone: PTB-024A19.F.  
BAC Library clone: PTB-024A19.F.  
Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
1 (sites)  
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.  
BAC end sequences of Library PTB  
Unpublished  
2 (bases 1 to 657)  
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.  
Direct Submission

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: chimpes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
Clones are derived from the chimpanzee AC library PTB This BAC end  
was generated during the R&D process and may have higher chance of  
clone tracking errors.

PRIMERS  
Sequencing: -21M13  
LIBRARY  
Vector : pKS145  
R.Site 1 : SacI  
R.Site 2 : SacI  
Location/Qualifiers  
I. .657  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="PTB-024A19.F"

## FEATURES

Source  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="PTB-024A19.F"

```

/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      151 a 156 c 186 g 164 t
ORIGIN

Query Match      84.0%; Score 21; DB 12; Length 657;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 acaaggtccttaatcgatcc 25
|||||
Db 362 ACAAGGTCCTTAATCGCATCC 342

RESULT 2
AG126122/c
LOCUS
DEFINITION Pan troglodytes DNA, clone: PTB-136J09.R, genomic survey sequence.
ACCESSION AG126122
VERSION AG126122.1 GI:16655287
KEYWORDS GSS; GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male BAC Library clone:PTB-136J09.R.
ORGANISM Pan troglodytes
REFERENCE
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimbges@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1. 688
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-136J09.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
1 others
BASE COUNT      148 a 174 c 187 g 178 t
ORIGIN

Query Match      84.0%; Score 21; DB 12; Length 688;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 acaaggtccttaatcgatcc 25
|||||
Db 349 ACAAGGTCCTTAATCGCATCC 329

RESULT 3
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      151 a 156 c 186 g 164 t
ORIGIN

Query Match      84.0%; Score 21; DB 12; Length 657;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 acaaggtccttaatcgatcc 25
|||||
Db 362 ACAAGGTCCTTAATCGCATCC 342

RESULT 4
AG190162/c
LOCUS
DEFINITION mt95f04.r1 Soares mouse 3NBMS Mus musculus cdna clone IMAGE:637663 5', mRNA sequence.
ACCESSION AA190162
VERSION AA190162.1 GI:1776747
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

```

```

BH222496/c
LOCUS
DEFINITION 1006107F04.y1 1006 - RescueMu Grid G Zea mays genomic, DNA sequence.
ACCESSION BH222496
VERSION BH222496.1 GI:16817401
KEYWORDS GSS.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 255)
AUTHORS Walbot,V.
TITLE Maize genomic sequences found using engineered RescueMu transposon
JOURNAL Unpublished (2001)
COMMENT Department of Biological Sciences
Stanford University
855 California Ave. Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Possible ligation site of ends cut by 2 different endonucleases.
Reverse complemented post-ligation sequence from source sequence.
Plate: 1006107 row: 12
Class: transposon-tagged.
Location/Qualifiers
1. 255
/organism="Zea mays"
/cultivar="mixed background W23/Al88/B73"
/db_xref="taxon:4577"
/clone_lib="1006 - RescueMu Grid G"
/tissue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Organ: leaf; Vector: RescueMu (engineered from pBluescript backbone); Site_1: BamHI; Site_2: BglII; RescueMu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on RescueMu, go to the web site 'www.zmdb.iastate.edu' and follow the links for 'RescueMu.' Grid G was grown at Stanford in 2000. DNA was extracted from leaf punches, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."
BASE COUNT      62 a 62 c 65 g 66 t
ORIGIN

Query Match      76.8%; Score 19.2; IH 12; Length 255;
Best Local Similarity 87.5%; Pred. No. 86;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 agagacaagtccttaatcgatcc 24
|||||
Db 220 ATAGACAAAGTCCTTAATCGCATC 197

RESULT 4
AG190162/c
LOCUS
DEFINITION mt95f04.r1 Soares mouse 3NBMS Mus musculus cdna clone IMAGE:637663 5', mRNA sequence.
ACCESSION AA190162
VERSION AA190162.1 GI:1776747
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

```

Unpublished (1999)  
Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
This clone is available royalty-  
free from the IMAGE Consortium (info@image.llnwd.net)  
This clone was previously sequenced  
from the 3' end  
High quality sequence stop: 456.

```
/sex= male
/tissue_type="Spleen"
/dev_stage="4 weeks"
```

/lab\_host="DH109"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5'-  
 TATTCAATCTGAATGGAGCGCGCGCTTTTTTTTTTTTTTTTTT  
 3']; double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. RNA  
 provided by Dr. Bertrand Jordan. Library went through  
 three rounds of normalization, and was constructed by  
 Bento Soares and M. Fatima Bonaldo."

Qy 2 gaacaaagqtccttaatgcacatcc 25  
 ||||| ||||| ||| |||||  
 Db 357 GAGACAAAGTCCTTACTCGAATCC 380

RESULT 6  
AI509341/C

LOCUS	A1509341
DEFINITION	mt195f04.y1 Soares mouse 5' mRNA sequence.
ACCESSION	A1509341
VERSION	A1509341.1 GI:4408246
KEYWORDS	EST.
SOURCE	house mouse.

KEYWORDS	
SOURCE	house mouse.
EST.	Mus musculus
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. REFERENCE: 1 (bases 1 to 507)

**AUTHORS**

Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurt, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.

Email: [Mouseest@watson.wustl.edu](mailto:Mouseest@watson.wustl.edu)  
This clone is available royalty-free through LILN: contact the

IMAGE Consortium (info@image.llnl.gov) for further information.  
This read is a RESEQUENCE of a previously sequenced mouse clone  
correct orientation)

Seq primer: -40RP from Gibco  
High quality sequence stop: 471  
POLYA-No.

# FEATURES

Location/Qualifiers

1. 507  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:637663"  
/clone\_lib="Soares mouse 3NBMS"  
/sex="male"  
/tissue\_type="Spleen"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTTCACCAATCTGAAGTGGAGCGCGCGCTGTGTGTGTGTGTGTGTGT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. RNA  
provided by Dr. Hertrand Jordan. Library went through  
three rounds of normalization, and was constructed by  
Bento Soares and M.Fatima Bonaldo."

131 a 89 c 99 g 188 t

# BASE COUNT

ORIGIN

Query Match 76.8%; Score 19.2; DB 9; Length 507;

Best Local Similarity 87.5%; Pred. NO. 1e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gagcaaggtccttaatcgcatcc 25

||||| ||||| |||||

Db 389 GAGCAAGAGTCTTACTCGAATCC 366

# RESULT 7

LOCUS

DEFINITION  
BJ187698 593 bp mRNA linear EST 24-JAN-2002  
BJ187698 normalized full length cDNA library, chloronemata,  
caulonemata and malformed buds Physcomitrella patens subsp. patens  
cDNA clone pphb41c18 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE  
1 (bases 1 to 593)  
Fujita,T., Shin-I,T., Seki,M., Kamiya,A., Uchiyama,I., Nishiyama,T.,  
Carninci,P., Hayashizaki,Y., Shinozaki,K., Kohara,Y. and Hasebe  
M.

TITLE  
Comparison of the moss Physcomitrella patens genome with flowering  
plants genome

JOURNAL

COMMENT

Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp

A backbone of the vector is pBluescript II, that was in vivo  
excised from a modified IPS phage vector (Mo bi Tec, Germany). XhoI  
digested-5' end of cDNA is ligated to SalI site of the vector, and  
the BamHI digested-3' end including poly-A tail is ligated to BamHI

site of the vector. cDNA insert could be amplified with  
conventional T7 and T3 primers. This normalized full-length cDNA  
library was generated basically according to the method described  
in Genome Research 10, 1617-1630 (2000), Carninci, P. et al.  
Protonemata were blended by the POLYTRON, and then cultivated on  
the BCD medium containing 0.5uM BA (benzylaminopurine) for 8 to 13  
days under the continuous light.

# FEATURES

Source

1. 593  
/organism="Physcomitrella patens subsp. patens"  
/db\_xref="taxon:145481"  
/clone="pphb41c18"  
/clone\_lib="normalized full length cDNA library,  
chloronemata, caulonemata and malformed buds"  
/tissue\_type="mixture of chloronemata, caulonemata and  
malformed buds"

BASE COUNT 133 a 139 c 175 g 146 t

# ORIGIN

Query Match 75.2%; Score 18.8; DB 10; Length 593;

Best Local Similarity 90.9%; Pred. NO. 1.e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 gacaaaggtccttaatcgcatcc 25

||||| ||||| |||||

Db 585 GACAAAGGTCTTAATCAGATAC 564

# RESULT 8

LOCUS

DEFINITION  
BJ170162 631 bp mRNA linear EST 24-JAN-2002  
BJ170162 full length cDNA library, chloronemata and young  
gametophores Physcomitrella patens subsp. patens cDNA clone  
pph24ml6 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE  
1 (bases 1 to 631)  
Fujita,T., Shin-I,T., Seki,M., Kamiya,A., Uchiyama,I., Nishiyama,T.,  
Carninci,P., Hayashizaki,Y., Shinozaki,K., Kohara,Y. and Hasebe  
M.

TITLE  
Comparison of the moss Physcomitrella patens genome with flowering  
plants genome

JOURNAL

COMMENT

Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp

A backbone of the vector is pBluescript(KS), that  
was in vivo excised from a modified IPS phage vector (Mo bi Tec,  
Germany). 5' end of the cDNA that was digested with XhoI was  
ligated to SalI site of the vector and the 3' end including polyA  
tail was ligated to BamHI site of the vector. cDNA insert could be  
amplified with conventional T7 and T3 primers. This full-length  
cDNA library was generated basically according to the method  
described in The Plant J 15, 707-720 (1998) Seki M. et al.  
Protonemata were blended by the POLYTRON, and then cultivated on  
the BCDATG medium for 13-14 days under the continuous light.

# FEATURES

Source

1. 631  
/organism="Physcomitrella patens subsp. patens"  
/db\_xref="taxon:145481"  
/clone="pph24ml6"  
/clone\_lib="full length cDNA library, chloronemata and  
young gametophores"

/tissue type="mixture of chloronemata and young gametophores with 2 to 5 leaves"

BASE COUNT 158 a 153 c 150 g 167 t 3 others  
ORIGIN

Query Match 75.2%; Score 18.8; DB 10; Length 631;

Best Local Similarity 90.9%; Pred. No. 1.7e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 gacaaagtcttaatacgcattcc 25

||||| ||||| ||||| |||||

DB 540 GACAAGTCTCTTAATCACATAC 561

RESULT 9

BI894401/C

LOCUS

DEFINITION ga54c11.y1 Moss EST library PPU Physcomitrella patens cDNA clone, PEP\_SOURCE\_ID:PPU080621 5' similar to TR:Q95827 Q95827 SUBMITTER: DEHYDROGENASE IRON-PROTEIN SUBUNIT ;, mRNA sequence.

ACCESSION BI894401

VERSION BI894401.1 GI:16137537

KEYWORDS EST.

SOURCE Physcomitrella patens.

ORGANISM

REFERENCE Physcomitrella patens.

AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.

Quatrano, R., Bashardes, S., Cove, D., Cuming, A., Knight, C., Clifton

, S., Marra, M., Hillier, L., Pape, D., Martin, J., Wylie, T., Underwood

, K., Theisinger, B., Allen, M., Bowers, Y., Person, B., Swaller, T.,

Stephens, M., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.

, Waterston, R., and Wilson, R.

Leeds/Wash U Moss EST Project

Unpublished (1999)

CONTACT: Ralph Quatrano

Leeds/Wash U Moss EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

Libraries were constructed by Dr. Stavros Bashardes as part of the

Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and

Washington Univ. in St. Louis (USA) DNA sequencing by: Washington

University Genome Sequencing Center For information on obtaining a

clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)

Seq primer: -40RP from Gibco

High quality sequence stop: 427.

Location/Qualifiers

1. .674

/organism="Physcomitrella patens"

/db\_xref="taxon:3218"

/clone="PEP\_SOURCE\_ID:PPU080621"

/clone\_lib="Moss EST library PPU"

/tissue\_type="protonemata: 7 day old tissue

anionium-grown"

/lab\_host="DH108"

/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:

XhoI; Construction of the cDNA library was carried out

using Strategenes 'UnizAP - cDNA synthesis kit'. cDNA

was constructed using an oligo dt primer/linker that

contains a XhoI site within it. Following ds cDNA

synthesis, EcoRI adapters were ligated to the blunt ends

and sample was digested with XhoI. The result is cDNA

with an EcoRI sticky end on one side and a XhoI sticky

end on the other. This cDNA was ligated directionally in

UnizAP arms. The vector is designed containing the

pBluescript sequence as well as lambda DNA and cDNA is

cloned within this pBluescript sequence. The vector was

then packaged using Gold gigacloning extracts. Library

was grown in XLBlue MRF' cells and amplified. The library

was excised by mass excision using Strategenes 'Mass excision kit' that uses exsist as a helper phage that releases the pBluescript sequence and circularises it as single stranded plasmids that are then packaged (by helper phage) and secreted out of the host cell as phagemids.

SOLR cells were transformed with phagemids and the library was plated out on LB-amp plates to select for transformants. Approximately 1,000,000 colonies were grown and recovered. The double stranded plasmid library was

recovered by using Quiaagen Midi prep kit. 2 micro grams of each library were used to transform DH10B cells by electroporation."

BASE COUNT 163 a 176 c 175 g 159 t 1 others  
ORIGIN

Query Match

Best Local Similarity 75.2%; Score 18.8; DB 10; Length 674;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 gacaaagtcttaatacgcattcc 25

||||| ||||| ||||| |||||

DB 425 GACAAGTCTCTTAATCACATAC 404

RESULT 10

BI186945/C

LOCUS

DEFINITION AR046G081.N05K06S Porcine Lymph node cDNA library, Cot 5 Sus scrofa

ACCESSION BI186945

VERSION BI186945.1 GI:15414895

KEYWORDS EST.

SOURCE pig.

ORGANISM

Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE 1 (bases 1 to 365)

AUTHORS Rink, A., Santschi, E. M., and Beattie, C. W.

TITLES Amplified, Normalized cDNA Libraries from a Porcine Model of

Orthopedic Implant Associated Staphylococcus aureus Infection

Unpublished (2001)

JOURNAL

COMMENT

Contact: Rink A

Department of Animal Biotechnology

College of Agriculture, Biotechnology and Natural Resources,

University of Nevada, Reno

MS 202, FA 103, 1664 N Virginia St, Reno, NV 89557-0236, USA

Tel: 775 784 1705

Fax: 775 784 1375

Email: arink@cabnr.unr.edu

Tissues and cells are derived from a porcine model for

implant-associated infection using 1000 cfu of Staphylococcus

aureus in a tibial transection, reduced and internally fixed with a

dynamic compression plate. NOTE: The sequences contain a 'cDNA

adapter' between the EcoRI site and the start of the EST. The

adapter sequence is 'AATTCGCGACGAG'.

Location/Qualifiers

1. .365

/organism="Sus scrofa"

/strain="crossbreed"

/db\_xref="taxon:9823"

/clone\_lib="Porcine Lymph node cDNA library, Cot 5"

/tissue\_type="Lymph node"

/cell\_type="mixed"

/dev\_stage="control, 5 month old castrated male"

/lab\_host="SOLR"

/note="Vector: pBSK; Site\_1: Eco RI; Site\_2: XhoI; Tissues

and cells are derived from a porcine model for

implant-associated infection using 1000 cfu of

Staphylococcus aureus in a tibial transection, reduced and

internally fixed with a dynamic compression plate. NOTE:

The sequences contain a 'cDNA adapter' between the EcoRI

site and the start of the EST. The adapter sequence is



us-09-623-329-3.rst

Query Match	74.4%	Score 18.6;	DB 12;	Length 601;
Best Local Similarity	84.0%;	Pred. No. 2.1e+02;		
Matches 21;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;
Qy	1	agadacaagggtcccttaatcgcatcc	25	

Db 593 AGACACAAGGCTCATTTATTGCAATCC 569  
 |||||  
 RESULT 13  
 BF430776

LOCUS 704 bp mRNA linear EST 20-FEB-2001  
 DEFINITION OG04E09T3 OG Oryza sativa cDNA clone OG04E09 5', mRNA sequence.  
 ACCESSION BF430776  
 VERSION BF430776.1 GI:11442877  
 KEYWORDS EST.  
 SOURCE Oryza sativa.  
 ORGANISM Eukaryota; Oryza sativa

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzeae; Oryza.

1 (bases 1 to 704)  
 Cushman, J.C. and Nomura, M.  
 A cDNA library from Oryza sativa cv. Pokkali leaves enriched for  
 salt-responsive genes  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Cushman JC  
 Department of Biochemistry  
 University of Nevada  
 MS200, Reno, NV 89557-0014, USA  
 Tel: 775-784-1918  
 Fax: 775-784-1650  
 Email: jcushman@unr.edu

PCR Primers  
 FORWARD: T7  
 BACKWARD: T3  
 Plate: OG04 row: E column: 9  
 Seq primer: T3  
 High quality sequence stop: 350  
 POLYA-No.

## FEATURES

source  
 1..704  
 /organism="Oryza sativa"  
 /cultivar="Pokkali"  
 /db\_xref="taxon:4530"  
 /clone="OG04E09"  
 /clone\_lib="OG"  
 /tissue\_type="leaf"  
 /dev\_stage="7 day-old (treated with 150 mM NaCl for 1, 2  
 and 7 days)."  
 /lab\_host="Escherichia coli"  
 /note="Vector: UNI-ZAP XR, pBluescriptSK- (Stratagene);  
 Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 155 a 182 c 170 g 197 t  
 ORIGIN

Query Match 74.4%; Score 18.6; DB 10; Length 704;  
 Best Local Similarity 84.0%; Pred. No. 2.1e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 aagacaaaggtccttaatcgcatcc 25

Db 589 AGTACAAAGGCTTATTCGCTTGC 613  
 |||||

RESULT 14  
 BF464363/c  
 LOCUS 773 bp mRNA linear EST 21-AUG-2001  
 DEFINITION 603204015F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:5269826 5',  
 mRNA sequence.

ACCESSION BF464363  
 VERSION BF464363.1 GI:15255019  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 773)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs@mail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM11680 row: n column: 03  
 High quality sequence stop: 764.

## FEATURES

source

1..773  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5269826"  
 /clone\_lib="NIH\_MGC\_97"  
 /lab\_host="DH10B"  
 /note="Organ: testis; Vector: pBluescriptR (modified  
 pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag  
 ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
 size-selected for average insert size 2.2 kb and  
 normalized to ROT 5. This is a primary library enriched  
 for full-length clones and constructed using the  
 Cap-trapper method (Carninci, in preparation). Library  
 constructed by M. Brownstein (NIH/NHGRI, National  
 Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 178 a 221 c 201 g 173 t  
 ORIGIN

Query Match 74.4%; Score 18.6; DB 10; Length 773;  
 Best Local Similarity 84.0%; Pred. No. 2.2e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 aagacaaaggtccttaatcgcatcc 25

Db 601 AGGACAAGGCTCTCAATCTCCTCC 577  
 |||||

RESULT 15

LOCUS W91495

DEFINITION m91495.1 Soares mouse embryo NbME13.5 14.5 Mus musculus cDNA  
 clone IMAGE:424074 5' similar to gb:M14043 ANNEXIN II (HUMAN);  
 qb:D10024 Mouse mRNA for protein-tyrosine kinase substrate p36  
 (MOUSE);, mRNA sequence.

ACCESSION W91495

VERSION W91495.1 GI:1407327

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 218)

AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.

Unpublished (1996)

CONTACT: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:258626  
Trace considered overall poor quality  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 1.  
Location/Qualifiers

FEATURES

source

1. .218  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:424074"  
/clone\_lib="Soares mouse embryo NIMEL3.5 14.5"  
/sex="unknown"  
/tissue\_type="embryo"  
/dev\_stage="13.5-14.5dpc total fetus"  
/lab\_host="DH10B"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5'  
TGTACCAATCTGAAGTGGAGCGCGCGGAAATTTTTTTTTTTTTTTTTT  
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2  
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne  
State Univ., from 2 ]; double-stranded cDNA was ligated to  
Eco RI adaptors (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of the modified  
pT7T3 vector. Library went through one round of  
normalization, and was constructed by Bento Soares and  
M.Fatima Bonaldo. "

BASE COUNT 53 a 54 c 62 g 49 t  
ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 218;  
Best Local Similarity 87.0%; Pred. No. 2.5e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 gagacaaggctcctaatacgcac 24  
|||||  
Db 56 GAGACAAGGCTCCTTAGAATC 78

Search completed: May 18, 2002, 17:22:05  
Job time: 14676 sec

CC the sequence provided in Fig 1 of the specification).  
 XX  
 SQ Sequence 1092 BP; 266 A; 236 C; 393 G; 197 T; 0 other;

Query Match 100.0%; Score 20; DB 20; Length 1092;  
 Best Local Similarity 100.0%; Pred. No. 0.59;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctttacaacctaaagc 20  
 |||||||||||||||||||

Db 714 ctccctttacaacctaaagc 733

## RESULT 4

AAT15396  
 ID AAT15396 standard; DNA; 1212 BP.

XX  
 AC AAT15396;

DT 19-JUL-1996 (first entry)

XX EBV nuclear antigen protein 1 gene.

DE EBV; nuclear antigen protein 1; EBNA1; immunoassay; baculovirus;  
 KW ds.

XX Epstein-Barr virus.  
 XX WO9602563-A1.

PN 01-FEB-1996.

XX 13-JUL-1995; 95WO-US08700.

XX 13-JUL-1994; 94US-0275614.

XX (CORR ) CORNELL RES FOUND INC.  
 XX O'Donnell ME;

XX WPI; 1996-105853/11.  
 DR P-PSDB; AAR88587.

XX Recovery of Epstein Barr Virus nuclear antigen protein 1 (EBNA1) -  
 PT useful in immunoassay system for detecting Epstein Barr Virus in  
 PT serum samples

PS Claim 12; Page 49; 82pp; English.

XX An isolated DNA (AAT15396) codes for the Epstein-Barr virus (EBV)  
 CC nuclear antigen protein 1 (EBNA1) (AAR88587), the viral encoded  
 CC protein which binds the latent phase origin (oriP) of EBV.  
 CC A recombinant DNA expression system comprises an expression  
 CC vector carrying an insert of heterologous DNA encoding EBNA-1.  
 CC The host cells may be e.g. Escherichia coli (see also AAT15397) or  
 CC Sf-9 insect cells (baculovirus expression system). The EBNA1 is  
 CC recovered from the host cell nucleus and used in EBV immunoassays.

XX Sequence 1212 BP; 307 A; 266 C; 429 G; 210 T; 0 other;

Query Match 100.0%; Score 20; DB 17; Length 1212;  
 Best Local Similarity 100.0%; Pred. No. 0.6;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctttacaacctaaagc 20  
 |||||||||||||||||||

Db 831 ctccctttacaacctaaagc 850

## RESULT 5

AAX03475

ID AAX03475 standard; DNA; 1748 BP.

XX  
 AC AAX03475;

DT 26-APR-1999 (first entry)

XX Epstein-Barr virus compact replicon.

DE EBV; replicon; episomal replication; OriP; EBNA1;

XX Epstein-Barr nuclear antigen 1; gene therapy; vector; ss.

XX Chimeric - Epstein-Barr virus.

OS Chimeric - Herpes simplex virus type 1.

XX Key Location/Qualifiers  
 FH misc\_feature 1..495

FT /\*tag= a  
 FT /note= "OriP, specifically claimed in Claim 3"

FT promoter

FT 496..616  
 FT /note= "herpes simplex virus 1 thymidine kinase  
 gene modified promoter"

FT CDS 627..1721

FT /\*tag= b

FT /product= Epstein-Barr nuclear antigen 1

FT /note= "EBNA1, specifically claimed in Claim 4"

XX WO9857658-A1.

XX 23-DEC-1998.

XX 14-APR-1998; 98WO-US07641.

XX 19-JUN-1997; 97US-0050206.

XX (GEO ) GEN HOSPITAL CORP.

XX Seed B;

XX WPI; 1999-080933/07.

XX Epstein-Barr virus replicons which support episomal replication -  
 PT comprise an OriP sequence and an EBNA1 sequence operably linked to a  
 PT promoter

XX Disclosure; Fig 1; 12pp; English.

XX This 1748 bp fragment acts as a compact Epstein-Barr virus (EBV)  
 CC replicon. Extensive deletion and mutagenesis studies showed that  
 CC it is possible to embed the cis- and trans-acting functions  
 CC necessary for EBV episomal replication into a fragment of less than  
 CC 2 kb. The 1748 bp fragment contains all sequences needed for  
 CC efficient expression of the Epstein-Barr nuclear antigen 1 (EBNA-1)  
 CC protein, with the exception of a polyA consensus sequence. It is  
 CC designed to be inserted upstream of a bidirectional polyA sequence  
 CC (see also AAX03476) in an appropriate vector. Plasmid vectors based  
 CC on this sequence replicate as episomes in the nucleus of transfected  
 CC mammalian cells. Compact EBV replicons find use in gene therapy  
 CC vectors, e.g. in gene delivery vehicles such as expression vectors.

XX Sequence 1748 BP; 453 A; 382 C; 546 G; 367 T; 0 other;

Query Match 100.0%; Score 20; DB 20; Length 1748;  
 Best Local Similarity 100.0%; Pred. No. 0.62;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctttacaacctaaagc 20  
 |||||||||||||||||||

Db 1340 ctccctttacaacctaaagc 1359

## RESULT 6

AAX09024

ID AAX90924 standard; DNA; 1925 BP.  
XX AAX90924;  
XX  
XX 17-JAN-2000 (first entry)  
XX  
XX Epstein Barr Virus Nuclear Antigen 1 (EBNA 1) DNA.  
XX  
XX Epstein Barr Virus Nuclear Antigen 1 DNA; plasmid pCMVEBNA; EBNA 1;  
KW episome; transfection; origin of replication; EBV orip; receptor;  
KW eukaryotic host cell; recombinant cell line; ion channel; gene therapy;  
KW multiple gene expression; transporter protein; transcription factor;  
KW adhesion molecule; antisense therapy; gene amplification;  
KW cell immortalisation; ds.  
XX  
XX Epstein-barr virus.  
XX  
XX  
XX Key Location/Qualifiers  
XX CDS 1..1925  
XX /\*tag= a  
XX /product= "EBNA 1"  
XX /trans\_except= (pos:799..800, aa:Gly)  
XX /note= "The sequence is described throughout the  
XX specification as being 1926 nucleotides long, but a  
XX sequence of only 1925 bp has been given in figure 2"  
XX  
XX WO9947647-A1.  
XX  
XX 23-SEP-1999.  
XX  
XX 12-FEB-1999; 99WO-US03307.  
XX  
XX 18-MAR-1998; 98US-0040961.  
XX 06-AUG-1998; 98US-0130114.  
XX  
XX (PHAR-) PHARMACOPEIA INC.  
XX  
XX Damaj BB, Horlick RA, Robbins AK;  
XX  
XX WPI: 1999-610610/52.  
XX P-PSDB; AAY28843.  
XX  
XX New method for expressing genes from recombinant eukaryotic cells,  
XX useful for gene therapy -  
XX  
XX Claim 24; Fig 2; 86pp; English.  
XX  
XX The present sequence is a DNA encoding Epstein Barr Virus Nuclear  
XX Antigen 1 (EBNA 1), which is obtained from commercially available  
XX plasmid pCMVEBNA. EBNA 1 protein is used to stably maintain episomes  
XX containing EBV origin of replication (oriP) and a gene encoding  
XX protein or RNA of interest. Eukaryotic host cells expressing EBNA 1  
XX cell lines are transfected with these episomes to produce recombinant  
XX cell lines expressing multiple genes of interest. This provides a  
XX rapid and reliable method of stably expressing multiple genes in  
XX transfected cells. The episomes are useful in the transfection of genes  
XX encoding receptors, transporter proteins, ion channels, adhesion  
XX molecules and transcription factors. The episomes carrying desired genes  
XX can also be used to transfect cells in gene therapy, antisense therapy,  
XX for gene amplification, cell immortalisation, etc.  
XX  
XX Sequence 1925 BP; 486 A; 352 C; 872 G; 215 T; 0 other;  
XX

Query Match 100.0%; Score 20; DB 20; Length 1925;  
Best Local Similarity 100.0%; Pred. No. 0.63;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctttacaacctaaggc 20  
Db 1544 ctccctttacaacctaaggc 1563

RESULT 7  
ID AAA50254 standard; DNA; 1926 BP.  
XX AAA50254;  
XX  
XX 07-NOV-2000 (first entry)  
XX  
XX Epstein Barr virus nuclear antigen 1 protein (EBNA1) DNA.  
XX  
XX EBV; nuclear antigen 1; EBNA1; episome; transfection; selection;  
KW gene therapy; ds.  
XX  
XX Epstein-barr virus.  
XX  
XX WO200047778-A1.  
XX  
XX 17-AUG-2000.  
XX  
XX 11-FEB-2000; 2000WO-US03547.  
XX  
XX 11-FEB-1999; 99US-0249585.  
XX (PHAR-) PHARMACOPEIA INC.  
XX  
XX Horlick RA, Chelsky D;  
XX  
XX WPI: 2000-515062/46.  
XX P-PSDB; AAY95856.  
XX  
XX Stably transfecting eukaryotic cells with at least one episome for the  
XX production of a desired protein in vitro and for gene therapy -  
XX  
XX Disclosure; Fig 2; 53pp; English.  
XX  
XX The present sequence is that of DNA encoding the Epstein-Barr virus  
XX (EBV) nuclear antigen protein 1 (EBNA1, see AAY95856). EBNA1 is  
XX utilised in a novel method for obtaining a eukaryotic cell that is  
XX stably transfected with at least one episome. This method involves  
XX transfecting a eukaryotic cell with: (1) a first episome comprising  
XX an EBV origin of replication (oriP, see AAA50253), a gene encoding a  
XX first protein whose expression results in cell death and a  
XX selectable marker for eukaryotic cells; and (2) a second episome  
XX comprising an EBV orip and a gene encoding a second protein, where  
XX expression of the second protein prohibits the occurrence of cell  
XX death resulting from expression of the first protein to produce  
XX doubly transfected cells which also express an antigen that  
XX promotes retention of the episomes by the cells. The doubly  
XX transfected cells are maintained under conditions in which the  
XX first and second proteins and the selectable marker are expressed,  
XX and the selective pressure specified by the marker is maintained.  
XX Under these conditions, only cells containing both episomes live.  
XX Preferably, EBNA1 is expressed from 1 of the episomes, and the  
XX protein of interest from the other episome. Either or both  
XX epitopes may further comprise a nucleic acid sequence encoding a  
XX protein desired to be expressed in the cell (e.g. a therapeutic  
XX protein), a nucleic acid encoding an RNA that is not intended to  
XX be translated (e.g. a therapeutic RNA), or a DNA sequence used as  
XX a tag for the cells. The method is applicable to cell culture or  
XX intact organisms, for gene therapy. It allows the rapid  
XX establishment of eukaryotic cells that stably and reliably express  
XX a gene of interest, using a novel method of selection, and  
XX maintenance of that selection without the need for exogenous  
XX selection factors, such as antibiotics.  
XX  
XX Sequence 1926 BP; 487 A; 352 C; 872 G; 215 T; 0 other;  
XX

Query Match 100.0%; Score 20; DB 21; Length 1926;  
Best Local Similarity 100.0%; Pred. No. 0.63;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctttacaacctaaggc 20

```
Db 1545 ctccctttacacctaaggc 1564
|||||
RESULT 8
AAF82902
ID AAF82902 standard; DNA; 1926 BP.
XX
AC AAF82902;
XX
DT 29-JUN-2001 (first entry)
XX
DE EBV tethering protein EBNA1 encoding DNA.
XX
KW Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;
KW Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;
KW EBV; latency-associated nuclear antigen; LANA; EBNA1; ds.
XX
OS Epstein-barr virus.
XX
FH Key Location/Qualifiers
FT 1..1926
FT CDS /*lag= a
XX
PN WO200125484-A2.
XX
PD 12-APR-2001.
XX
PF 29-SEP-2000; 2000WO-US26908.
XX
PR 01-OCT-1999; 99US-0410399.
XX
PA (UNMI ) UNIV MICHIGAN.
XX
PI Robertson ES, Cotter MA;
XX
DR WPI: 2001-281736/29.
XX
DR P-PSDB; AAB62332.
XX
PT A composition for use in gene therapy comprises an expression vector
PT that includes a nucleic acid sequence encoding a nucleic acid binding
PT protein -
XX
PS Disclosure; Fig 9C; 60pp; English.
XX
CC The invention provides a composition comprising nucleic acid, histone H1
CC protein and expression vector operationally encoding a protein suitable
CC for tethering the nucleic acid to the histone H1 protein, where the
CC tethering protein is LANA. The composition is useful in aiding the
CC retention of the viral DNA in the host cell. The viral vector encodes a
CC protein suitable for tethering DNA to histone H1. Methods for screening
CC for compounds which are agonistic or antagonistic for the tethering of
CC viral proteins to histone H1 and DNA binding sites are useful for
CC developing the method of viral transfer. The composition has applications
CC to gene therapy, including the treatment of multiple sclerosis,
CC Parkinson's disease, Huntington disease and diabetes. The present
CC sequence represents the nucleotide sequence of the Epstein-barr virus
CC (EBV) tethering protein ENNA1.
XX
SQ Sequence 1926 BP; 487 A; 352 C; 872 G; 215 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 1926;
Best Local Similarity 100.0%; Pred. No. 0.63;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctttacacctaaggc 20
|||||
Db 1545 ctccctttacacctaaggc 1564

RESULT 9
AAF75454
ID AAX90923 standard; DNA; 5452 BP.
XX
AC AAX90923;
XX
DT 17-JAN-2000 (first entry)
XX
DE Anti-sense strand of pCMVEBNA plasmid.
XX
KW Anti-sense strand; plasmid pCMVEBNA; EBNA 1; episome; transfection;
KW Epstein Barr Virus Nuclear Antigen 1; origin of replication;
KW EBV orip; eucaryotic host cell; recombinant cell line; ion channel;
KW multiple gene expression; receptor; transporter protein; gene therapy;
KW transcription factor; adhesion molecule; antisense therapy;
KW gene amplification; cell immortalisation; ss.
XX
```

OS Epstein-barr virus.  
OS Cytomegalovirus.  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT complement (3032..4957)  
FT /tag= a  
FT /product= "EBNA 1 protein"  
FT /note= "Epstein Barr Virus Nuclear Antigen 1"  
XX  
XX  
PN WO9947647-A1.  
XX  
XX  
PD 23-SEP-1999.  
XX  
XX  
PF 12-FEB-1999; 99WO-US03307.  
XX  
PR 18-MAR-1998; 98US-0040961.  
PR 06-AUG-1998; 98US-0130114.  
XX  
XX (PHAR-) PHARMACOEIA INC.  
XX  
XX Dama] BB, Horlick RA, Robbins AK;  
XX  
XX WPI: 1999-610610/52.  
XX  
PT New method for expressing genes from recombinant eukaryotic cells,  
PT useful for gene therapy  
XX  
XX Example 1: Fig 1; 86pp; English.  
XX  
XX The present sequence is an anti-sense strand of commercially available  
CC plasmid pCMV-EBNA from which Epstein Barr Virus Nuclear Antigen 1 (EBNA  
CC 1) encoding DNA is obtained. EBNA 1 protein is used to stably maintain  
CC episomes containing EBV origin of replication (oriP) and a gene encoding  
CC protein or RNA of interest. Eucaryotic host cells expressing EBNA 1  
CC protein are transfected with these episomes to produce recombinant  
CC cell lines expressing multiple genes of interest. This provides a  
CC rapid and reliable method of stably expressing multiple genes in  
CC transfected cells. The episomes are useful in the transfection of genes  
CC encoding receptors, transporter proteins, ion channels, adhesion  
CC molecules and transcription factors. The episomes carrying desired genes  
CC can also be used to transfect cells in gene therapy, antisense therapy,  
CC for gene amplification, cell immortalisation, etc.  
XX  
SQ Sequence 5452 BP; 1108 A; 1736 C; 1246 G; 1362 T; 0 other;

Query Match 100.0%; Score 20; DH 20; Length 5452;  
Best Local Similarity 100.0%; Pred. No. 0.7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccctttacaactaagc 20  
Db 877 CTCCTTTACACCTAAGGC 858  
|||||  
|||||

RESULT 11  
AAQ40324  
ID AAQ40324 standard; cDNA; 8575 BP.  
XX  
XX  
AC AAQ40324;  
XX  
XX 09-AUG-1993 (first entry)  
XX  
XX Sequence of pHEBO30.  
XX  
XX Neurotrophic factor; growth promoting activity; GPA; ss.  
XX  
XX Gallus domesticus.  
XX  
XX Key Location/Qualifiers  
FH misc\_feature 781..3061  
FT /tag= a

FT repeat\_region /label= orip  
FT 882..1474 /tag= b  
FT 2587..2665 /tag= c  
FT /label= dyad region  
FT 3062  
FT /tag= d  
FT /note= "HSV TK Term 3' end"  
FT 3159..3737  
FT /tag= e  
FT /label= CMV  
FT 3738..2901  
FT /tag= f  
FT 2902..3988  
FT /tag= g  
FT /label= SP6  
FT 3989..3993  
FT /tag= h  
FT /note= "SP6 KNA start"  
FT 4510..4646  
FT /tag= i  
FT /label= SV40 early poly A  
FT /note= "g"  
FT 4647..4691  
FT /tag= j  
FT /label= SV40 origin  
FT 4692  
FT /tag= k  
FT /label= HSV TK Term 5' end  
FT complement (5185..6207)  
FT /tag= l  
FT /label= HPG gene from PLC89-- hygroresistance  
FT complement (6234..6279)  
FT /tag= m  
FT /label= HSV TK  
FT complement (6306..6359)  
FT /tag= n  
FT complement (6360..6479)  
FT /tag= o  
FT complement (6480..6803)  
FT /tag= p  
FT /label= PBR 322; let-kpeptide  
FT complement (7604..8448)  
FT /tag= q  
FT /label= beta lastanase  
XX  
XX WO9307270-A.  
PN  
XX 15-AUG-1993.  
PD  
XX  
XX 29-SEP-1992; 92WO-US08258.  
PF  
XX 01-OCT-1991; 91US-0769622.  
PR  
XX (GETH) GENENTECH INC.  
PA (UYOR-) UNIV OREGON HEALTH SCI.  
XX  
XX Cachianes G, Eckenstein FP, Leung D, Nishi R;  
PI WPI: 1993-134459/16.  
XX  
XX Nucleic acid encoding growth promoting activity - used for  
PT developing prods. for diagnosis and treatment of neurological  
PT diseases and disorders  
XX  
XX Disclosure; Fig 4; 74pp; English.  
XX  
XX Eyes from embryonic day 15 (E15) chicken embryos were dissected and  
CC total RNA was isolated. The RNA was used to prepare cDNA which was  
CC than ligated to SfiI digested lambda-HEBO vector. The E15 cDNA  
CC library was screened using oligonucleotide probes based on the  
CC partial amino acid sequences of GPA obtd. by microsequencing of 3

CC different peptide fragments of purified GPA. The oligonucleotides  
 CC are referred to as o-GPA-1, o-GPA-2 and o-GPA-3. Lambda CE15 #19  
 CC hybridized with all three probes. The expression of GPA involves  
 CC the use of a plasmid expression vector (pHEB030) containing the oriP  
 CC region from EBV. pHEB030 comprises the strong CMV promoter, a multiple  
 CC cloning region for insertion of foreign (exogenous) genes downstream  
 CC of the CMV promoter, the oriP region of EBV for plasmid replication  
 CC in host cells expressing EBNA-1, a hygromycin resistance gene for  
 CC selection in eukaryotes, the origin of replication from pBR322 for  
 CC replication in prokaryotes, and an ampicillin resistance gene for  
 CC selection in prokaryotes. In the sequence "N" is used to  
 CC designate the nucleotides that comprise the arbitrary 550 base pair  
 CC CDNA insert in pHEB030.

XX SQ Sequence 8575 BP; 1989 A; 2167 G; 2032 G; 2047 T; 350 other;

Query Match 100.0%; Score 20; DB 14; Length 8575;  
 Best Local Similarity 100.0%; Pred. No. 0.73;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cccccctttacacctaagc 20  
 |||  
 Db 99 cccccctttacacctaagc 118

## RESULT 12

AAZ23778/c  
 ID AAZ23778 standard; DNA; 8705 BP.

XX AC AAZ23778;

DT 14-JAN-2000 (first entry)

DE Vector pShuttle DNA.

KW Antisense; DNA library; identification; multiple cloning site; MCS;  
 inhibition; ss.

XX OS Synthetic.

PN WO9950457-A1.

PD 07-OCT-1999.

PF 28-MAR-1999; 99WO-US06742.

PR 28-MAR-1998; 98US-0079792.

PR 06-NOV-1998; 98US-0107504.

PA (UTAH ) UNIV UTAH RFS FOUND.

PI Ruffner DE, Pierce ML, Chen Z;

XX DR WPI; 1999-610866/52.

PT Production of antisense libraries, used for identifying antisense  
 PT agents and for identifying target sites for antisense-mediated  
 PT inhibition of a selected gene.

XX PS Claim 16; Page 43-50; 63pp; English.

XX CC This invention describes a novel method for generating an antisense  
 CC library targeted to a selected RNA transcript. The methods can be used  
 CC for identifying antisense agents and for identifying target sites for  
 CC antisense-mediated inhibition of a selected gene. The use of a direct  
 CC library for target site selection significantly simplifies the screening  
 CC process, since only very small libraries need be prepared and assayed.  
 CC This sequence represents the vector pShuttle which is used in the method  
 CC of the invention.

XX SQ Sequence 8705 BP; 1875 A; 2575 C; 2078 G; 2177 T; 0 other;

Query Match 100.0%; Score 20; DB 20; Length 8705;  
 Best Local Similarity 100.0%; Pred. No. 0.73;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cccccctttacacctaagc 20  
 |||  
 Db 6742 CTCCTTTACACCTAAGGC 6723

## RESULT 13

AAV21683  
 ID AAV21683 standard; DNA; 9600 BP.

XX AC AAV21683;

DT 17-AUG-1998 (first entry)

DE Vector plasmid pCMVkmLTR-EPI.

XX Polynucleotide delivery; plasmid pCMVkmLTR-EPI; vector;  
 KW gene therapy; vaccine; polycationic agent; ss.

XX OS Chimeric - Epstein-Barr virus.

OS Chimeric - Adeno-associated virus.

OS Chimeric - Cytomegalovirus.

OS Chimeric - Bos taurus.

XX Key Location/Qualifiers

FT CDS 14..2594

FT /tag- a

FT /product- "EBV nuclear antigen A"

FT /misc\_feature 2623..4559

FT /tag- b

FT /note- "EBV origin of replication"

FT repeat\_unit 4928..5104

FT /tag- c

FT /rpt\_type- INVERTED

FT /note- "AAV inverted terminal repeat"

FT repeat\_unit 7189..7355

FT /tag- d

FT /rpt\_type- INVERTED

FT /note- "AAV inverted terminal repeat"

FT promoter 5112..6734

FT /tag- e

FT /note- "CMV immediate-early enhancer/promoter"

FT terminator 6818..7050

FT /tag- f

FT /note- "bovine growth hormone polyA sequence"

XX WO9806437-A2.

XX 19-FEB-1998.

XX 13-AUG-1997; 97WO-US14465.

XX 13-AUG-1996; 96US-0023867.

XX (CHIR ) CHIRON CORP.

XX Cohen F, Dubois-Stringfellow N, Dworki V, Innis MA;  
 PI Murphy JE, Tetsuo U, Zukermann R;

XX WPI; 1998-159296/14.

XX Polycationic agents based on alpha-amino acids, able to complex  
 PT with nucleic acid - to facilitate its entry into cell, condense it  
 PT and protect it against serum degradation, particularly for use in  
 PT gene therapy

XX PS Disclosure; Page 77-80; 100pp; English.

XX CC This polynucleotide comprises the DNA sequence of vector plasmid



CC PCMVkNTR-EPI, which contains an Epstein-Barr virus (EBV) origin  
 CC of replication from plasmid pCEP4, a coding region for EBV nuclear  
 CC antigen A from pCEP4, a pair of inverted terminal repeats from  
 CC adeno-associated virus, a cytomegalovirus enhancer/promoter, a  
 CC bovine growth hormone polyA sequence, and a kanamycin resistance  
 CC selectable marker. Polynucleotides encoding polypeptides, such as  
 CC erythropoietin or leptin, and ribozymes and antisense  
 CC polynucleotides can be inserted into the vector. The vector is  
 CC preferred for use in novel compositions and methods for improved  
 CC polynucleotide delivery into cells. In these methods, polycationic  
 CC agents are used to increase the frequency of uptake of a  
 CC nucleic acid (see also AA21684-86) into a cell. The polycationic  
 CC agent can condense with the nucleic acid and inhibit serum and/or  
 CC nuclease degradation of the nucleic acid. The nucleic acid can be  
 CC a vector, may express a therapeutic protein or a vaccinating viral  
 CC or cancer antigen, or is itself therapeutic (antisense or  
 CC ribozyme). The methods and compositions can be used in the gene  
 CC therapy of many diseases.

XX SQ Sequence 9600 BP; 2326 A; 2376 C; 2817 G; 2081 T; 0 other;

Query Match 100.0%; Score 20; DB 19; Length 9600;  
 Best Local Similarity 100.0%; Pred. No. 0.74;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctttacacctaaggc 20  
 Db 1974 ctccctttacacctaaggc 1993

## RESULT 14

AAAS3873  
 ID AAAS3873 standard; DNA; 9725 BP.

XX AC AAAS3873;

XX 03-JAN-2001 (first entry)

XX DE Expression vector pRIG19.

XX KW Vector; endogenous gene; activation; over-expression;  
 KW erythropoietin; growth hormone; drug discovery;  
 KW granulocyte colony stimulating factor; ds.

XX OS Synthetic.

XX PN WO200049162-A2.

XX PD 24-AUG-2000.

XX PF 22-FEB-2000; 2000WO-US04429.

XX PR 19-FEB-1999; 99US-0253022.

XX PR 08-MAR-1999; 99US-0263814.

XX PR 26-MAR-1999; 99US-0276820.

XX PA (ATHE-) ATHERSYS INC.

XX PI Harrington JJ, Sherf B, Rundlett S;

XX DR WPI; 2000-549276/50.

XX Non-targeted activation of endogenous genes, e.g. for the production of  
 PT erythropoietin, growth hormone or granulocyte-colony stimulating factor  
 PT proteins and for drug discovery

XX PS Disclosure; Fig 30; 240pp; English.

XX CC New methods, vectors and cells are described for non-targeted  
 CC activation and over-expression of endogenous genes. The vector  
 CC constructs comprise transcriptional regulatory sequences (TRS) and  
 CC unpaired splice donor sequences (USDS), preferably the vectors

CC comprise (in sequential order) a TRS, an USDS, a rare cutting  
 CC restriction site (RCRS) and a linearization site (LS) with a second  
 CC TRS linked to a selectable marker (SM) lacking a polyadenylation  
 CC signal. The methods, vectors and cells comprising the vectors may  
 CC be used for the non-targeted activation and over-expression of  
 CC endogenous genes, e.g. for the production of proteins (including  
 CC erythropoietin, growth hormone or granulocyte-colony stimulating  
 CC factor) and drug discovery. The advantage of these methods are that  
 CC endogenous genes including those associated with human disease and  
 CC development, may be activated and isolated without prior knowledge  
 CC of the sequence structure, function or expression profile of  
 CC the genes being known.

XX SQ Sequence 9725 BP; 2425 A; 2507 C; 2448 G; 2343 T; 2 other;

Query Match 100.0%; Score 20; DB 21; Length 9725;  
 Best Local Similarity 100.0%; Pred. No. 0.74;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctttacacctaaggc 20  
 Db 3196 ctccctttacacctaaggc 3215

## RESULT 15

AAAS3879  
 ID AAAS3879 standard; DNA; 9732 BP.

XX AC AAAS3879;

XX 03-JAN-2001 (first entry)

XX DE Expression vector pRIG-TP.

XX KW Vector; endogenous gene; activation; over-expression;  
 KW erythropoietin; growth hormone; drug discovery;  
 KW granulocyte colony stimulating factor; ds.

XX OS Synthetic.

XX PN WO200049162-A2.

XX PD 24-AUG-2000.

XX PF 22-FEB-2000; 2000WO-US04429.

XX PR 19-FEB-1999; 99US-0253022.

XX PR 08-MAR-1999; 99US-0263814.

XX PR 26-MAR-1999; 99US-0276820.

XX PA (ATHE-) ATHERSYS INC.

XX PI Harrington JJ, Sherf B, Rundlett S;

XX DR WPI; 2000-549276/50.

XX Non-targeted activation of endogenous genes, e.g. for the production of  
 PT erythropoietin, growth hormone or granulocyte-colony stimulating factor  
 PT proteins and for drug discovery

XX PS Example 15; Fig 37; 240pp; English.

XX CC New methods, vectors and cells are described for non-targeted  
 CC activation and over-expression of endogenous genes. The vector  
 CC constructs comprise transcriptional regulatory sequences (TRS) and  
 CC unpaired splice donor sequences (USDS), preferably the vectors  
 CC comprise (in sequential order) a TRS, an USDS, a rare cutting  
 CC restriction site (RCRS) and a linearization site (LS) with a second  
 CC TRS linked to a selectable marker (SM) lacking a polyadenylation  
 CC signal. The methods, vectors and cells comprising the vectors may  
 CC be used for the non-targeted activation and over-expression of  
 CC endogenous genes, e.g. for the production of proteins (including

CC erythropoietin, growth hormone or granulocyte-colony stimulating  
 CC factor) and drug discovery. The advantage of these methods are that  
 CC endogenous genes including those associated with human disease and  
 CC development, may be activated and isolated without prior knowledge  
 CC of the sequence structure, function or expression profile of  
 CC the genes being known.

XX

SQ Sequence 9732 BP; 2427 A; 2512 C; 2443 G; 2348 T; 2 other;

Query Match 100.0%; Score 20; DB 21; Length 9732;

Best Local Similarity 100.0%; Pred. No. 0.74; 0; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccctttacaaccttaagc 20

|||||

Db 3203 ctccctttacaaccttaagc 3222

Search completed: May 18, 2002, 17:45:49  
 Job time: 13738 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:45:49 ; Search time 1024.22 Seconds  
(without alignments)  
41.908 Million cell updates/sec

Title: US-09-623-329-3

Perfect score: 25

Sequence: 1 agagacaaggctcctaatacgcaccc 25

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SID55/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID55/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID55/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID55/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID55/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID55/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID55/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID55/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID55/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID55/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID55/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID55/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID55/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID55/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID55/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID55/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	AAZ11653	Oligo specific for
2	21	84.0	36	AAQ52597	EBV target sequenc
3	21	84.0	624	AAT15397	PCR-generated eEBN
4	21	84.0	1092	AAQ03478	Epstein-Barr nucle
5	21	84.0	1212	AAT15396	EBV nuclear antige
6	21	84.0	1748	AAQ03475	Epstein-Barr virus
7	21	84.0	1925	AAQ90924	Epstein Barr virus
8	21	84.0	1926	AAA50254	Epstein Barr virus
9	21	84.0	1926	AAF82902	EBV tethering prot

c 10	21	84.0	2580	21	AAA75454	Nucleotide sequenc
c 11	21	84.0	5452	20	AAQ90923	Anti-sense strand
c 12	21	84.0	8575	14	AAQ40324	Sequence of pHEB03
c 13	21	84.0	8705	20	AAZ23778	vector pshuttle DN
c 14	21	84.0	9600	19	AAV21683	vector plasmid pCM
c 15	21	84.0	9725	21	AAV53873	Expression vector
c 16	21	84.0	9732	21	AAV53879	Expression vector
c 17	21	84.0	9738	21	AAV53874	Expression vector
c 18	21	84.0	9873	21	AAV53875	Expression vector
c 19	21	84.0	10054	21	AAV53876	Expression vector
c 20	21	84.0	10380	20	AAZ22248	Nucleotide sequenc
c 21	21	84.0	10596	14	AAQ51731	Plasmid pCISBON f
c 22	21	84.0	10596	17	AAT40348	Plasmid pCISBON f
c 23	21	84.0	10596	20	AAV15650	Nucleotide sequenc
c 24	21	84.0	11265	19	AAV59501	Plasmid PREP7:CTL
c 25	21	84.0	16080	21	AAV59553	DNA clone pCEK C1.
c 26	17.6	70.4	4615	19	AAV05285	Saccharomyces cere
c 27	17.4	69.6	22977	22	AAK77120	Human immune haema
c 28	17.2	68.8	19031	20	AAV13104	Enterococcus faeca
c 29	17	68.0	637	21	AAF09555	Fusarium venenatum
c 30	17	68.0	706	23	AAV83203	DNA encoding novel
c 31	17	68.0	1053	18	AAV24972	H. pylori inner me
c 32	17	68.0	1153	19	AAV14036	H. pylori GHPO 102
c 33	17	68.0	1459	21	AAV44863	Arabidopsis thalia
c 34	17	68.0	2287	21	AAV88670	Kat novel nng2 gen
c 35	17	68.0	2341	22	AAV44683	Human full-length
c 36	17	68.0	2533	21	AAV15835	Human prostate can
c 37	17	68.0	2575	22	AAV44852	Human contig polyn
c 38	16.8	67.2	142	22	AAK81532	Human immune haema
c 39	16.8	67.2	146	19	AAV36484	Partial nucleotide
c 40	16.8	67.2	295	21	AAV41337	Human secreted exp
c 41	16.8	67.2	442	21	AAV02606	Human secreted pro
c 42	16.8	67.2	2013	21	AAV51624	Arabidopsis thalia
c 43	16.8	67.2	2365	19	AAV11890	A. thaliana acetol
c 44	16.8	67.2	2365	19	AAV11891	A. thaliana acetol
c 45	16.8	67.2	2365	22	AAV06570	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1

AAZ11653

ID AAZ11653 standard; DNA; 25 BP.

XX AAZ11653;

XX 19-NOV-1999 (first entry)

XX Oligo specific for Epstein Barr Nuclear Antigen 1 (EBNA-1) RNA.

XX Epstein Barr Virus; EBV infection; viral; gene transcription; EBEP-1;

XX Epstein Barr early RNA; Epstein Barr nuclear antigen 1; EBNA-1; LMP-1;

XX latent membrane protein; LMP-2; vIL10; BCRF-1; BARF1; BDLF2; NASBA;

XX EBV-associated malignancy; primer; ss.

XX Synthetic.

XX Epstein-barr virus.

XX WO945155-A2.

XX 10-SEP-1999.

XX 01-MAR-1999; 99WO-EP01392.

XX 04-MAR-1998; 98EP-0200655.

XX 14-DEC-1998; 98EP-0204231.

XX (ALKU ) AKZO NOBEL NV.

XX Vervoort MBHJ, Van Den Brule AJC, Middeldorp JM;

XX WPI; 1999-551051/46.

```
XX Identifying Epstein Barr Virus infection -
PT Claim 11; Page 20; 50pp; English.
PS
XX The invention provides methods for identifying an Epstein Barr Virus
CC (EBV) infection, that comprises determining viral gene transcription
CC patterns by amplification of specific RNA sequences. The binding sites
CC of the oligos suitable for amplification are located in the following
CC genes: Epstein Barr early RNA (EBER-1), Epstein Barr nuclear antigen 1
CC (EBNA-1), latent membrane protein 1 (LMP-1), LMP-2, and vIL10 (BCRF-1),
CC BARF1 and BDLF2. The method comprises (a) amplifying a target sequence
CC within one or more RNA(s) transcribed from above gene sequences and the
CC (b) detecting the amplified products, determining the transcription
CC pattern and identifying the corresponding EBV-associated malignancy. The
CC RNA is amplified using a transcription based amplification technique
CC such as NASBA. The invention is used to diagnose malignant and
CC non-malignant EBV-associated diseases. Sequences AA211651-54 represent
CC oligos specific for EBNA-1 RNA.
XX
SQ Sequence 25 BP; 8 A; 7 C; 5 G; 5 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.004;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agagacaaaggtcttaatgcgcatcc 25
   |||||||
DB 1 agagacaaaggtcttaatgcgcatcc 25

RESULT 2
AAQ52597/c
ID AAQ52597 standard; RNA; 36 BP.
XX
AC AAQ52597;
XX
DT 26-MAY-1994 (first entry)
XX
DE EBV target sequence 21.
XX
XX RNA; enzyme; enzymatic RNA molecule; ERM; cleave; RNA; mRNA; HnRNA;
KW picornavirus; HIV; immunodeficiency virus; hepatitis B virus; HBV;
KW papilloma virus; HPV; Epstein-Barr virus; EBV; TClv;
KW T-cell leukaemia virus; hepatitis C virus; HCV; cytomegalovirus;
KW influenza virus; HSV; herpes simplex virus; vector; immune response;
KW antibody; ribozyme; viral RNA; treatment; ss.
XX
OS Synthetic.
XX
XX WO9323569-A.
XX
XX 25-NOV-1993.
XX
XX 29-APR-1993; 93WO-US04020.
XX
XX 11-MAY-1992; 92US-0882689.
PR 14-MAY-1992; 92US-0882712.
PR 14-MAY-1992; 92US-0882713.
PR 14-MAY-1992; 92US-0882714.
PR 14-MAY-1992; 92US-0882715.
PR 14-MAY-1992; 92US-0882823.
PR 14-MAY-1992; 92US-0882824.
PR 14-MAY-1992; 92US-0882886.
PR 14-MAY-1992; 92US-0882888.
PR 14-MAY-1992; 92US-0882889.
PR 14-MAY-1992; 92US-0882921.
PR 14-MAY-1992; 92US-0883823.
PR 14-MAY-1992; 92US-0883849.
PR 14-MAY-1992; 92US-0884073.
PR 14-MAY-1992; 92US-0884074.
PR 14-MAY-1992; 92US-0884333.
PR 14-MAY-1992; 92US-0884422.

14-MAY-1992; 92US-0884431.
PR 14-MAY-1992; 92US-0884436.
PR 14-MAY-1992; 92US-0884521.
PR 31-JUL-1992; 92US-0923738.
PR 26-AUG-1992; 92US-0936086.
PR 18-SEP-1992; 92US-0948359.
PR 15-OCT-1992; 92US-0963322.
PR 07-DEC-1992; 92US-0987129.
PR 07-DEC-1992; 92US-0987130.
PR 07-DEC-1992; 92US-0987133.
XX
PA (RIBO-) RIBOZYME PHARM INC.
XX
XX Draper KG, Dudycz LW, Holecek JJ, Macejak DG, Mamane JA;
PI McSwiggen JA;
XX
XX WPI; 1993-386599/48.
XX
XX Enzymatic RNA molecules - used to inhibit viral replication,
PT infection and gene expression
XX
XX Claim 5; Fig 10; 287pp; English.
XX
XX The sequences (AAQ52577-601) are pref. Epstein-Barr virus target
CC sequences for enzymatic RNA molecules. The RNA molecules are
CC complementary to a substrate binding region in the specified gene
CC target. They also have enzymatic activity, in that they specifically
CC cleave RNA in the target. The ERMs interfere with viral replication and
CC therefore have anti-viral properties. They can be used to attenuate
CC viruses to be used in vaccines.
XX
SQ Sequence 36 BP; 12 A; 6 C; 10 G; 8 U; 0 other;

Query Match 84.0%; Score 21; DB 14; Length 36;
Best Local Similarity 100.0%; Pred. No. 0.44;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaagggccttaatgcgcatcc 25
   |||||||
DB 24 ACAAGGTCCTTAATGCATCC 4

RESULT 3
AAT15397/c
ID AAT15397 standard; DNA; 624 BP.
XX
AC AAT15397;
XX
DT 19-JUL-1996 (first entry)
XX
DE PCR-generated eEBNA1 protein gene.
XX
KW EBV; nuclear antigen protein 1; EBNA1; Immunoassay; ds.
XX
OS Epstein-barr virus.
XX
XX Key Location/Qualifiers
FT primer_bind complement (1..24)
FT /*tag= a
FT /*note= "primer for eEBNA1"
FT primer_bind 584..624
FT /*tag= b
FT /*note= "primer for eEBNA1"
XX
XX WO9602563-A1.
XX
XX 01-FEB-1996.
XX
XX 13-JUL-1995; 95WO-US08700.
XX
XX 13-JUL-1994; 94US-0275614.
XX
```

(CORR ) CORNELL RES FOUND INC.

O'Donnell ME;

WPI; 1996-105853/11.

P-PSDB; AAR88588.

Recovery of Epstein Barr Virus nuclear antigen protein 1 (EBNA1) -  
useful in immunoassay system for detecting Epstein Barr Virus in  
serum samples

Example 22; Page 51; 82pp; English.

A DNA fragment (AAT15397) was generated by PCR amplification of  
the Epstein-Barr virus (EBV) nuclear antigen protein 1 (EBNA1)  
gene (see also AAT15396) in plasmid p291 and NdeI/BamHI digestion of  
the PCR product. It was ligated into pET3c and the resulting  
plasmid was used to transform Escherichia coli strain BL21  
(DE3)pLysS cells. The recombinant EBNA1, designated eBNAL  
(AAR88588), can be recovered from the nuclei of host cells for use  
in EBV immunoassays.

Sequence 624 BP; 148 A; 132 C; 195 G; 149 T; 0 other;

Query Match 84.0%; Score 21; DB 17; Length 624;

Best Local Similarity 100.0%; Pred. No. 0.76;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaaggtccttaacgcaccc 25

|||||

Db 425 ACAAGGTCTTAATCGCATCC 405

RESULT 4

AAX03478/C

ID AAX03478 standard; DNA; 1092 BP.

AC AAX03478;

26-APR-1999 (first entry)

Epstein-Barr nuclear antigen 1 (EBNA1) coding region.

EBV; replicon; episomal replication; OriP; EBNA1;

Epstein-Barr nuclear antigen 1; gene therapy; vector; ss.

Epstein-Barr virus.

WO9857658-A1.

23-DEC-1998.

14-APR-1998; 98WO-US07641.

19-JUN-1997; 97US-0050206.

(GEHO ) GEN HOSPITAL CORP.

Seed B;

WPI; 1999-080933/07.

Epstein-Barr virus replicons which support episomal replication -  
comprise an OriP sequence and an EBNA1 sequence operably linked to a  
promoter

Claim 4; Page -; 12pp; English.

This is the Epstein-Barr nuclear antigen 1 (EBNA1) coding region.  
A novel, claimed nucleic acid sequence which supports episomal  
replication in a mammalian cell comprises an Epstein-Barr virus  
(EBV) OriP sequence (see AAX03477) and an EBNA1 sequence operably

CC linked to a promoter, the whole sequence having a length of less  
CC than 3 kb, and preferably less than 1.8 kb (see AAX03475). The  
CC nucleic acid sequence preferably further comprises a bidirectional  
CC polyadenylation consensus (see AAX03476), and is designed to be  
CC inserted upstream of such a polyadenylation sequence in an  
CC appropriate vector. The claimed nucleic acid sequence acts as a  
CC compact EBV replicon. Plasmid vectors based on this sequence  
CC replicate as episomes in the nucleus of transfected mammalian  
CC cells. Compact EBV replicons find use in gene therapy vectors,  
CC e.g. in gene delivery vehicles such as expression vectors.  
CC (NB, the sequence of EBNA1 was created with respect to Claim 4 and  
CC the sequence provided in Fig 1 of the specification).

XX Sequence 1092 BP; 266 A; 236 C; 393 G; 197 T; 0 other;

Query Match 84.0%; Score 21; DB 20; Length 1092;

Best Local Similarity 100.0%; Pred. No. 0.84;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaaggtccttaacgcaccc 25

|||||

Db 917 ACAAGGTCTTAATCGCATCC 897

RESULT 5

AAT15396/C

ID AAT15396 standard; DNA; 1212 BP.

AC AAT15396;

19-JUL-1996 (first entry)

EBV nuclear antigen protein 1 gene.

EBV; nuclear antigen protein 1; EBNA1; immunoassay; baculovirus;

ds.

Epstein-barr virus.

WO9602563-A1.

01-FEB-1996.

13-JUL-1995; 95WO-US08700.

13-JUL-1994; 94US-0275614.

(CORR ) CORNELL RES FOUND INC.

O'Donnell ME;

WPI; 1996-105853/11.

P-PSDB; AAR88587.

Recovery of Epstein Barr Virus nuclear antigen protein 1 (EBNA1) -  
useful in immunoassay system for detecting Epstein Barr Virus in  
serum samples

Claim 12; Page 49; 82pp; English.

An isolated DNA (AAT15396) codes for the Epstein-Barr virus (EBV)  
nuclear antigen protein 1 (EBNA1) (AAR88587), the viral encoded  
protein which binds the latent phase origin (oriP) of EBV.

A recombinant DNA expression system comprises an expression  
vector carrying an insert of heterologous DNA encoding EBNA-1.  
The host cells may be e.g. Escherichia coli (see also AAT15397) or  
Sf-9 insect cells (baculovirus expression system). The EBNA1 is  
CC recovered from the host cell nucleus and used in EBV immunoassays.

XX Sequence 1212 BP; 307 A; 266 C; 429 G; 210 T; 0 other;

Query Match 84.0%; Score 21; DB 17; Length 1212;  
Best Local Similarity 100.0%; Pred. No. 0.86;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaaggtccttaatgcgcacc 25  
|||||  
DB 1034 ACAAGGTCCTTAATGCGCATCC 1014

RESULT 6  
AA03475/c  
ID AAX03475 standard; DNA; 1748 BP.  
XX AC AAX03475;  
XX DT 26-APR-1999 (first entry)  
XX DE Epstein-Barr virus compact replicon.  
XX KW EBV; replicon; episomal replication; OriP; EBNA1;  
XX KW Epstein-Barr nuclear antigen 1; gene therapy; vector; ss.  
XX OS Chimeric - Epstein-Barr virus.  
XX OS Chimeric - Herpes simplex virus type 1.  
XX FH Key Location/Qualifiers  
FT misc\_feature 1..495  
FT /\*tag= a  
FT /note= "OriP, specifically claimed in Claim 3"  
FT promoter 496..616  
FT /note= "herpes simplex virus 1 thymidine kinase  
FT gene modified promoter"  
FT CDS 627..1721  
FT /\*tag= b  
FT /product= Epstein-Barr nuclear antigen 1  
FT /note= "EBNA1, specifically claimed in Claim 4"  
XX PN WO9857658-A1.  
XX DT 23-DEC-1998.  
XX PF 14-APR-1998; 98WO-US07641.  
XX PR 19-JUN-1997; 97US-0050206.  
XX PA (GENO ) GEN HOSPITAL CORP.  
XX PI Seed B;  
XX WPI; 1999-080933/07.  
XX DT Epstein-Barr virus replicons which support episomal replication  
XX PT comprise an OriP sequence and an EBNA1 sequence operably linked to a  
XX PT promoter  
XX PS Disclosure; Fig 1; 12pp; English.  
XX CC This 1748 bp fragment acts as a compact Epstein-Barr virus (EBV)  
XX CC replicon. Extensive deletion and mutagenesis studies showed that  
XX CC it is possible to embed the cis- and trans-acting functions  
XX CC necessary for EBV episomal replication into a fragment of less than  
XX CC 2 kb. The 1748 bp fragment contains all sequences needed for  
XX CC efficient expression of the Epstein-Barr nuclear antigen 1 (EBNA-1)  
XX CC protein, with the exception of a polyA consensus sequence. It is  
XX CC designed to be inserted upstream of a bidirectional polyA sequence  
XX CC (see also AAX03476) in an appropriate vector. Plasmid vectors based  
XX CC on this sequence replicate as episomes in the nucleus of transfected  
XX CC mammalian cells. Compact EBV replicons find use in gene therapy  
XX CC vectors, e.g. in gene delivery vehicles such as expression vectors.  
XX SQ Sequence 1748 BP; 453 A; 382 C; 546 G; 367 T; 0 other;

Query Match 84.0%; Score 21; DB 20; Length 1748;  
Best Local Similarity 100.0%; Pred. No. 0.92;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaaggtccttaatgcgcacc 25  
|||||  
DB 1543 ACAAGGTCCTTAATGCGCATCC 1523

RESULT 7  
AAX90924/c  
ID AAX90924 standard; DNA; 1925 BP.  
XX AC AAX90924;  
XX DT 17-JAN-2000 (first entry)  
XX DE Epstein Barr Virus Nuclear Antigen 1 (EBNA 1) DNA.  
XX KW Epstein Barr Virus Nuclear Antigen 1 DNA; plasmid pcMVEBNA; EBNA 1;  
XX KW episome; transfection; origin of replication; EBV OriP; receptor;  
XX KW eucaryotic host cell; recombinant cell line; ion channel; gene therapy;  
XX KW multiple gene expression; transporter protein; transcription factor;  
XX KW adhesion molecule; antisense therapy; gene amplification;  
XX KW cell immortalisation; ds.  
XX KW Epstein-Barr virus.  
XX OS Key Location/Qualifiers  
FT CDS 1..1925  
FT /\*tag= a  
FT /product= "EBNA 1"  
FT /transl\_except= (pos:799..800, aa:Gly)  
FT /note= "The sequence is described throughout the  
FT specification as being 1926 nucleotides long, but a  
FT sequence of only 1925 bp has been given in figure 2"  
XX PN WO9947647-A1.  
XX DT 23-SEP-1999.  
XX PF 12-FEB-1999; 99WO-US03307.  
XX PR 18-MAR-1998; 98US-0040961.  
XX PR 06-AUG-1998; 98US-0130114.  
XX PA (PHAR-) PHARMACOPETIA INC.  
XX PI Dama j HB, Horlick RA, Robbins AK;  
XX WPI; 1999-610610/52.  
XX DR P-PSDB; AAY28843.  
XX DT New method for expressing genes from recombinant eukaryotic cells,  
XX PT useful for gene therapy  
XX PS Claim 24; Fig 2; 86pp; English.  
XX CC The present sequence is a DNA encoding Epstein Barr Virus Nuclear  
XX CC Antigen 1 (EBNA 1), which is obtained from commercially available  
XX CC plasmid pcMVEBNA. EBNA 1 protein is used to stably maintain episomes  
XX CC containing EBV origin of replication (oriP) and a gene encoding  
XX CC protein or RNA of interest. Eucaryotic host cells expressing EBNA 1  
XX CC protein are transfected with these episomes to produce recombinant  
XX CC cell lines expressing multiple genes of interest. This provides a  
XX CC rapid and reliable method of stably expressing multiple genes in  
XX CC transfected cells. The episomes are useful in the transfection of genes  
XX CC encoding receptors, transporter proteins, ion channels, adhesion  
XX CC molecules and transcription factors. The episomes carrying desired genes  
XX CC can also be used to transfect cells in gene therapy, antisense therapy,  
XX CC for gene amplification, cell immortalisation, etc.  
XX SQ Sequence 1925 BP; 486 A; 352 C; 872 G; 215 T; 0 other;

Query Match 84.0%; Score 21; DB 20; Length 1925;  
Best Local Similarity 100.0%; Pred. No. 0.94;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaaggtccttaatgcaccc 25  
|||||  
DB 1747 ACAAGGTCTTAATGCATCC 1728

RESULT 8  
AAA50254/c  
ID AAA50254 standard; DNA; 1926 BP.  
XX AC AAA50254;  
XX DT 07-NOV-2000 (first entry)  
XX DE Epstein Barr virus nuclear antigen 1 protein (EBNA1) DNA.  
XX KW EBV; nuclear antigen 1; EBNA1; episome; transfection; selection;  
XX KW gene therapy; ds.  
XX OS Epstein-barr virus.  
XX PN WO200047778-A1.  
XX PD 17-AUG-2000.  
XX PF 11-FEB-2000; 2000WO-US03547.  
XX PR 11-FEB-1999; 99US-0249585.  
XX PH (PHAR-) PHARMACOPEDIA INC.  
XX PI Horlick RA, Chelsky D;  
XX WP: 2000-515062/46.  
XX P-PSDB; AAY95856.  
Stably transfecting eukaryotic cells with at least one episome for the production of a desired protein in vitro and for gene therapy -  
Disclosure; Fig 2; 53pp; English.

The present sequence is that of DNA encoding the Epstein-Barr virus (EBV) nuclear antigen protein 1 (EBNA1, see AAY95856). EBNA1 is utilised in a novel method for obtaining a eukaryotic cell that is stably transfected with at least one episome. This method involves transfecting a eukaryotic cell with: (1) a first episome comprising an EBV origin of replication (oriP, see AAA50253), a gene encoding a first protein whose expression results in cell death and a selectable marker for eukaryotic cells; and (2) a second episome comprising an EBV oriP and a gene encoding a second protein, where expression of the second protein prohibits the occurrence of cell death resulting from expression of the first protein to produce doubly transfected cells which also express an antigen that promotes retention of the episomes by the cells. The doubly transfected cells are maintained under conditions in which the first and second proteins and the selectable marker are expressed, and the selective pressure specified by the marker is maintained. Under these conditions, only cells containing both episomes live. Preferably, EBNA1 is expressed from 1 of the episomes, and the protein of interest from the other episome. Either or both episomes may further comprise a nucleic acid sequence encoding a protein desired to be expressed in the cell (e.g. a therapeutic protein), a nucleic acid encoding an RNA that is not intended to be translated (e.g. a therapeutic RNA), or a DNA sequence used as a tag for the cells. The method is applicable to cell culture or intact organisms, for gene therapy. It allows the rapid establishment of eukaryotic cells that stably and reliably express a gene of interest, using a novel method of selection, and

CC maintenance of that selection without the need for exogenous selection factors, such as antibiotics.

SQ Sequence 1926 BP; 487 A; 352 C; 872 G; 215 T; 0 other;

Query Match 84.0%; Score 21; DB 21; Length 1926;  
Best Local Similarity 100.0%; Pred. No. 0.94;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaaggtccttaatgcaccc 25  
|||||  
DB 1748 ACAAGGTCTTAATGCATCC 1728

RESULT 9  
AAF82902/c  
ID AAF82902 standard; DNA; 1926 BP.  
XX AC AAF82902;  
XX DT 29-JUN-2001 (first entry)  
XX DE EBV tethering protein EBNA1 encoding DNA.  
XX KW Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;  
XX KW Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;  
XX KW EBV; latency-associated nuclear antigen; LANA; EBNA1; ds.  
XX OS Epstein-barr virus.  
XX FH Key Location/Qualifiers  
XX CDS 1..1926  
XX FT /\*tag= a  
XX PN WO200125484-A2.  
XX PD 12-APR-2001.  
XX PF 29-SEP-2000; 2000WO-US26908.  
XX PR 01-OCT-1999; 99US-0410399.  
XX PA (UNMI ) UNIV MICHIGAN.  
XX PI Robertson ES, Cotter MA;  
XX WP: 2001-281736/29.  
XX P-PSDB; AAB62332.  
A composition for use in gene therapy comprises an expression vector that includes a nucleic acid sequence encoding a nucleic acid binding protein -  
Disclosure; Fig 9C; 60pp; English.  
The invention provides a composition comprising nucleic acid, histone H1 protein and expression vector operatively encoding a protein suitable for tethering the nucleic acid to the histone H1 protein, where the tethering protein is LANA. The composition is useful in aiding the retention of the viral DNA in the host cell. The viral vector encodes a protein suitable for tethering DNA to histone H1. Methods for screening for compounds which are agonistic or antagonistic for the tethering of viral proteins to histone H1 and DNA binding sites are useful for developing the method of viral transfer. The composition has applications to gene therapy, including the treatment of multiple sclerosis, Parkinson's disease, Huntington disease and diabetes. The present sequence represents the nucleotide sequence of the Epstein-barr virus (EBV) tethering protein ENNA1.  
SQ Sequence 1926 BP; 487 A; 352 C; 872 G; 215 T; 0 other;

```
Query Match      84.0%; Score 21; DB 22; Length 1926;
Best Local Similarity 100.0%; Pred. No. 0.94;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 acaaggtccttaatcgatcc 25
    |||||
Db 1748 ACAAGGTCCTTAATCGCATCC 1728

RESULT 10
ID AAA75454/C
XX AAA75454 standard; DNA; 2580 BP.
AC AAA75454;
XX
XX
XX 15-JAN-2001 (first entry)
XX
XX Nucleotide sequence of the Epstein Barr nuclear antigen.
XX
XX Origin of replication; protein-protein interaction; replication;
KW two-hybrid system; nuclear antigen; ss.
XX
XX Epstein-barr virus.
OS
XX
XX US6114111-A.
XX
XX 05-SEP-2000.
XX
XX 30-MAR-1998; 98US-0050863.
XX
XX 30-MAR-1998; 98US-0050863.
XX
XX (RICE-) RIGEL PHARM INC.
XX
XX Luo Y, Payan D, Huang B;
PI WPI; 2000-593546/56.
XX
XX Composition for detecting protein-protein interactions in a mammalian
PT two hybrid system comprises bait and test vector which consist of
PT selection gene, vector viral origin of replication and fusion gene
XX
XX Disclosure; Column 17-20; 18pp; English.
XX
XX The present sequence represents the Epstein Barr nuclear antigen.
XX It is used to produce bait vectors of the invention. The specification
XX describes a compositions and methods for a genetic system of detecting
XX protein-protein interactions in a mammalian host cell. The system
XX comprises bait and test, both containing selection genes, and viral
XX origin of replications which require bound viral replication proteins
XX to effect replication. The compositions is useful for detecting an
XX interaction between a bait protein and a test protein. It is useful in
XX a mammalian two-hybrid system for detecting protein-protein interactions
XX in a mammalian host cell.
XX
XX Sequence 2580 BP; 632 A; 512 C; 1054 G; 382 T; 0 other;

Query Match      84.0%; Score 21; DB 21; Length 2580;
Best Local Similarity 100.0%; Pred. No. 0.99;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 acaaggtccttaatcgatcc 25
    |||||
Db 2130 ACAAGGTCCTTAATCGCATCC 2110

RESULT 11
AXX90923
ID AAX90923 standard; DNA; 5452 BP.
XX
XX AAX90923;
XX
```

```
DT 17-JAN-2000 (first entry)
XX Anti-sense strand of pCMVEBNA plasmid.
DE
XX
XX Anti-sense strand; plasmid pCMVEBNA; EBNA 1; episome; transfection;
KW Epstein Barr Virus Nuclear Antigen 1; origin of replication;
KW EBV orip; eucaryotic host cell; recombinant cell line; ion channel;
KW multiple gene expression; receptor; transporter protein; gene therapy;
KW transcription factor; adhesion molecule; antisense therapy;
KW gene amplification; cell immortalisation; ss.
XX
XX Epstein-barr virus.
OS
XX Cytomegalovirus.
OS
XX Synthetic.
XX
XX Key Location/Qualifiers
FT complement (3032..4957)
FT /*tag= a
FT /product= "EBNA 1 protein"
FT /note= "Epstein Barr Virus Nuclear Antigen 1"
XX
XX WO9947647-A1.
XX
XX 23-SEP-1999.
XX
XX 12-FEB-1999; 99WO-US03307.
XX
XX 18-MAR-1998; 98US-0040961.
XX
XX 06-AUG-1998; 98US-0130114.
XX
XX (PHAR-) PHARMACOPETA INC.
XX
XX Damaj BB, Horlick RA, Robbins AK;
XX
XX WPI; 1999-610610/52.
XX
XX New method for expressing genes from recombinant eukaryotic cells,
XX useful for gene therapy
XX
XX Example 1; Fig 1; 86pp; English.
XX
XX The present sequence is an anti-sense strand of commercially available
XX plasmid pCMVEBNA from which Epstein Barr Virus Nuclear Antigen 1 (EBNA
XX 1) encoding DNA is obtained. EBNA 1 protein is used to stably maintain
XX episomes containing EBV origin of replication (orip) and a gene encoding
XX protein or RNA of interest. Eucaryotic host cells expressing EBNA 1
XX protein are transfected with these episomes to produce recombinant
XX cell lines expressing multiple genes of interest. This provides a
XX rapid and reliable method of stably expressing multiple genes in
XX transfected cells. The episomes are useful in the transfection of genes
XX encoding receptors, transporter proteins, ion channels, adhesion
XX molecules and transcription factors. The episomes carrying desired genes
XX can also be used to transfect cells in gene therapy, antisense therapy,
XX for gene amplification, cell immortalisation, etc.
XX
XX Sequence 5452 BP; 1108 A; 1736 C; 1246 G; 1362 T; 0 other;

Query Match      84.0%; Score 21; DB 20; Length 5452;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 acaaggtccttaatcgatcc 25
    |||||
Db 674 acaaggtccttaatcgatcc 694

RESULT 12
AAQ40324/C
ID AAQ40324 standard; CDNA; 8575 BP.
XX
XX AAQ40324;
XX
```



PT Nucleic acid encoding growth promoting activity - used for  
PT developing prods. for diagnosis and treatment of neurological  
PT diseases and disorders  
XX  
PS Disclosure; Fig 4; 74pp; English.  
XX  
CC Eyes from embryonic day 15 (E15) chicken embryos were dissected and  
CC total RNA was isolated. The RNA was used to prepare cDNA which was  
CC then ligated to SfiI digested lambda-HEBO vector. The E15 cDNA  
CC library was screened using oligonucleotide probes based on the  
CC partial amino acid sequences of GPA obtd. by microsequencing of 3  
CC different peptide fragments of purified GPA. The oligonucleotides  
CC are referred to as o-GPA-1, o-GPA-2 and o-GPA-3. Lambda CE15 #19  
CC hybridized with all three probes. The expression of GPA involves  
CC the use of a plasmid expression vector (PHEBO30) containing the orip  
CC region from EBV. PHEBO30 comprises the strong CMV promoter, a multiple  
CC cloning region for insertion of foreign (exogenous) genes downstream  
CC of the CMV promoter, the orip region of EBV for plasmid replication  
CC in host cells expressing EBNA-1, a hygromycin resistance gene for  
CC selection in eukaryotes, the origin of replication from pBR322 for  
CC replication in prokaryotes, and an ampicillin resistance gene for  
CC selection in prokaryotes. In the sequence "N" is used to  
CC designate the nucleotides that comprise the arbitrary 350 base pair  
CC cDNA insert in PHEBO30.  
XX  
SQ Sequence 8575 BP; 1989 A; 2167 C; 2032 G; 2037 T; 350 other;

PT Query Match 84.0%; Score 21; DB 14; Length 8575;  
PT Best Local Similarity 100.0%; Pred. No. 1-2;  
PT Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 5 acaaggtccttaatcgatcc 25  
Dd 302 ACAAGGTCCTTAATCGCATCC 282  
RESULT 13  
AAZ23778  
ID AAZ23778 standard; DNA; 8705 BP.  
XX  
AC AAZ23778;  
XX  
DT 14-JAN-2000 (first entry)  
XX  
DE Vector pShuttle DNA.  
XX  
KW Antisense; DNA library; identification; multiple cloning site; MCS;  
KW inhibition; ss.  
XX  
OS Synthetic.  
XX  
PN WO9950457-A1.  
XX  
PD 07-OCT-1999.  
XX  
PF 28-MAR-1999; 99WO-US06742.  
XX  
PR 28-MAR-1998; 98US-0079792.  
PR 06-NOV-1998; 98US-0107504.  
XX  
PA (UTAH ) UNIV UTAH RES FOUND.  
XX  
PI Ruffner DE, Pierce ML, Chen Z;  
XX  
DR WPI; 1999-610866/52.  
XX  
PT Production of antisense libraries, used for identifying antisense  
PT agents and for identifying target sites for antisense-mediated  
PT inhibition of a selected gene.  
XX  
PS Claim 16; Page 43-50; 63pp; English.  
XX

DT 09-AUG-1993 (first entry)  
XX  
DE Sequence of PHEBO30.  
XX  
KW Neurotrophic factor; growth promoting activity; GPA; ss.  
XX  
OS Gallus domesticus.  
FH Location/Qualifiers  
FT 781..3061  
FT /\*tag= a  
FT /label= orip  
FT repeat\_region 882..1474  
FT /\*tag= b  
FT misc\_feature 2587..2665  
FT /\*tag= c  
FT /label= dyad region  
FT misc\_feature 3062  
FT /\*tag= d  
FT /note= "HSV TK Term 3' end"  
FT promoter 3159..3737  
FT /\*tag= e  
FT /label= CMV  
FT misc\_RNA 3738..2901  
FT /\*tag= f  
FT promoter 2902..3988  
FT /\*tag= g  
FT /label= SP6  
FT misc\_RNA 3989..3993  
FT /\*tag= h  
FT /note= "SP6 RNA start"  
FT misc\_feature 4510..4646  
FT /\*tag= i  
FT /label= SV40 early poly A  
FT /\*tag= j  
FT /note= "9"  
FT misc\_feature 4647..4691  
FT /\*tag= k  
FT /label= SV40 origin  
FT misc\_feature 4692  
FT /\*tag= l  
FT /label= HSV TK Term 5' end  
FT complement (5185..6207)  
FT /\*tag= m  
FT /label= HPG gene from PLC89-- hygrosistance  
FT complement (6234..6279)  
FT /\*tag= n  
FT /label= HSV TK  
FT complement (6306..6359)  
FT /\*tag= o  
FT complement (6360..6479)  
FT /\*tag= p  
FT complement (6480..6803)  
FT /\*tag= q  
FT /label= PBR 322; tet-Rpeptide  
FT complement (7604..8448)  
FT /\*tag= r  
FT /label= beta lastanase  
XX  
PN WO9307270-A.  
XX  
PD 15-APR-1993.  
XX  
PF 29-SEP-1992; 92WO-US08258.  
XX  
PR 01-OCT-1991; 91US-0769622.  
XX  
PA (GETH ) GENENTECH INC.  
PA (UYOR-) UNIV OREGON HEALTH SCI.  
XX  
PI Cachianes G, Eckenstein FP, Leung D, Nishi R;  
XX  
DR WPI; 1993-134459/16.  
XX

CC This invention describes a novel method for generating an antisense  
CC library targeted to a selected RNA transcript. The methods can be used  
CC for identifying antisense agents and for identifying target sites for  
CC antisense-mediated inhibition of a selected gene. The use of a direct  
CC library for target site selection significantly simplifies the screening  
CC process, since only very small libraries need be prepared and assayed.  
CC This sequence represents the vector pShuttle which is used in the method  
CC of the invention.

XX Sequence 8705 BP; 1875 A; 2575 C; 2078 G; 2177 T; 0 other;

Query Match 84.0%; Score 21; DB 20; Length 8705;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaaggctcttaatcgatcc 25  
|||||

DB 6539 acaaggctcttaatcgatcc 6559

RESULT 14  
AAV21683/C  
ID AAV21683 standard; DNA; 9600 BP.

XX AC AAV21683;

XX DT 17-AUG-1998 (first entry)

XX DE Vector plasmid pCMVkmITR-EPI.

XX KW Polynucleotide delivery; plasmid pCMVkmITR-EPI; vector;  
KW gene therapy; vaccine; polycationic agent; ss.

XX OS Chimeric - Epstein-Barr virus.  
OS Chimeric - Adeno-associated virus.  
OS Chimeric - Cytomegalovirus.  
XX Chimeric - Bos taurus.

XX FH Key Location/Qualifiers  
FT CDS 14..2594

FT /tag= a

FT /product= "EBV nuclear antigen A"

FT misc\_feature 2623..4559

FT /tag= b

FT /note= "EBV origin of replication"

FT repeat\_unit 4528..5104

FT /tag= c

FT /rpt\_type= INVERTED

FT /note= "AAV inverted terminal repeat"

FT repeat\_unit 7189..7355

FT /tag= d

FT /rpt\_type= INVERTED

FT /note= "AAV inverted terminal repeat"

FT promoter 5112..6734

FT /tag= e

FT /note= "CMV immediate-early enhancer/promoter"

FT terminator 6818..7050

FT /tag= f

FT /note= "bovine growth hormone polyA sequence"

XX WO9806437-A2.

XX PN 19-FEB-1998.

XX PD 13-AUG-1997; 97WO-US14465.

XX PF 13-AUG-1996; 96US-0023867.

XX PR (CHIR ) CHIRON CORP.

XX PA Cohen F., Dubois-Stringfellow N., Dworki V., Innis MA;

XX PI Murphy JE, Tetsuo U, Zukermann R;

XX DR

XX WPI; 1998-159296/14.  
XX Polycationic agents based on alpha-amino acids, able to complex  
XX with nucleic acid - to facilitate its entry into cell, condense it  
XX and protect it against serum degradation, particularly for use in  
XX gene therapy  
XX  
XX Disclosure: Page 77-80; 100pp; English.

XX This polynucleotide comprises the DNA sequence of vector plasmid  
XX pCMVkmITR-EPI, which contains an Epstein-Barr virus (EBV) origin  
XX of replication from plasmid pCEP4, a coding region for EBV nuclear  
XX antigen A from pCEP4, a pair of inverted terminal repeats from  
XX adeno-associated virus, a cytomegalovirus enhancer/promoter, a  
XX bovine growth hormone polyA sequence, and a kanamycin resistance  
XX selectable marker. Polynucleotides encoding polypeptides, such as  
XX erythropoietin or leptin, and ribozymes and antisense  
XX polynucleotides can be inserted into the vector. The vector is  
XX preferred for use in novel compositions and methods for improved  
XX polynucleotide delivery into cells. In these methods, polycationic  
XX agents are used to increase the frequency of uptake of a  
XX nucleic acid (see also AAV21684-86) into a cell. The polycationic  
XX agent can condense with the nucleic acid and inhibit serum and/or  
XX nuclease degradation of the nucleic acid. The nucleic acid can be  
XX a vector, may express a therapeutic protein or a vaccinating viral  
XX or cancer antigen, or is itself therapeutic (antisense or  
XX ribozyme). The methods and compositions can be used in the gene  
XX therapy of many diseases.

XX Sequence 9600 BP; 2326 A; 2376 C; 2817 G; 2081 T; 0 other;

Query Match 84.0%; Score 21; DB 19; Length 9600;  
Best Local Similarity 100.0%; Pred. No. 1.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaaggctcttaatcgatcc 25

DB 2177 ACAAGGCTCTTAATCGATCC 2157

RESULT 15

AAA53873/C

ID AAA53873 standard; DNA; 9725 BP.

XX AC AAA53873;

XX DT 03-JAN-2001 (first entry)

XX DE Expression vector pRIG19.

XX KW Vector; endogenous gene; activation; over-expression;

XX KW erythropoietin; growth hormone; drug discovery;

XX KW granulocyte colony stimulating factor; ds.

XX OS Synthetic.

XX PN WO200049162-A2.

XX PD 24-AUG-2000.

XX PF 22-FEB-2000; 2000WO-US04429.

XX PR 19-FEB-1999; 99US-0253022.

XX PR 08-MAR-1999; 99US-0263814.

XX PR 26-MAR-1999; 99US-0276820.

XX PA (ATHE-) ATHERSYS INC.

XX PI Harrington JJ, Sherf B, Rundlett S;

XX WPI; 2000-549276/50.

**us-09-623-329-3.rng**

XX	Non-targeted activation of endogenous genes, e.g. for the production of
PT	erythropoietin, growth hormone or granulocyte-colony stimulating factor
PT	proteins and for drug discovery
XX	
XX	Disclosure: Fig 30: 240pp: English.
XX	
CC	New methods, vectors and cells are described for non-targeted
CC	activation and over-expression of endogenous genes. The vector
CC	constructs comprise transcriptional regulatory sequences (TRS) and
CC	unpaired splice donor sequences (USDS), preferably the vectors
CC	comprise (in sequential order) a TRS, an USDS, a rare cutting
CC	restriction site (RCRS) and a linearization site (LS) with a second
CC	TRS linked to a selectable marker (SM) lacking a polyadenylation
CC	signal. The methods, vectors and cells comprising the vectors may
CC	be used for the non-targeted activation and over-expression of
CC	endogenous genes, e.g. for the production of proteins (including
CC	erythropoietin, growth hormone or granulocyte-colony stimulating
CC	factor) and drug discovery. The advantage of these methods are that
CC	endogenous genes including those associated with human disease and
CC	development, may be activated and isolated without prior knowledge
CC	of the sequence structure, function or expression profile of
CC	the genes being known.
XX	
XX	Sequence 9725 BP: 2425 A; 2507 C; 2448 G; 2343 T; 2 other;
SO	

Query Match 84.0%; Score 21; DB 21; length 9725;  
Best Local Similarity 100.0%; Pred. No. 1.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 acaaggctccttaatcgcattcc 25  
|||||  
Db 3399 ACAAGGTCCTTAATCGCATCC 3379

Search completed: May 18, 2002, 17:45:51  
Job time: 13740 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 13:17:29 : Search time 8624.33 Seconds  
(without alignments)  
31.300 Million cell updates/sec

Title: us-09-623-329-2

Perfect score: 20

Sequence: 1 ctcctttacacctaagc 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estim:\*  
4: em\_estmd:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	657	12 AG045201	AG045201 Pan trogl
2	20	100.0	688	12 AG126122	AG126122 Pan trogl
3	17.4	87.0	653	9 BB621639	BB621639 BB621639
4	17.4	87.0	678	10 BE376997	BE376997 601227175
5	17	85.0	541	9 AU234961	AU234961 AU234961
6	16.8	84.0	420	10 BF441557	BF441557 257892 MA
7	16.8	84.0	606	12 AZ289149	AZ289149 RPCI-23-5
8	16.8	84.0	632	12 AZ787637	AZ787637 2M0034N09
9	16.8	84.0	891	10 BE271342	BE271342 601140504
10	16.8	84.0	904	10 BE917358	BE917358 601666260
11	16.8	84.0	975	10 BG259805	BG259805 602371951
12	16.4	82.0	644	12 AZ327575	AZ327575 1M0050M21
13	16.4	82.0	728	12 BH540661	BH540661 BCGWH69TF
14	16.4	82.0	734	10 BG965724	BG965724 602830560
15	16.4	82.0	801	10 BF341631	BF341631 602016082
16	15.8	79.0	293	10 BG224066	BG224066 1M0023F09
17	15.8	79.0	299	10 T10058	T10058 seq768 b4HB

18	15.8	79.0	312	9 AV333600	AV333600 AV333600
19	15.8	79.0	322	10 BI033413	BI033413 PMI-NN120
20	15.8	79.0	328	12 AQ090328	AQ090328 HS-3000.B
21	15.8	79.0	351	10 BF353713	BF353713 QV2-HT069
22	15.8	79.0	373	10 C00061	C00061 HUMGS000430
23	15.8	79.0	400	9 AA812675	AA812675 a179f11.s
24	15.8	79.0	420	9 AI087335	AI087335 oz45q12.x
25	15.8	79.0	423	9 AW318628	AW318628 un02612.y
26	15.8	79.0	427	12 A2248764	A2248764 RPCI-23-4
27	15.8	79.0	431	12 AQ056500	AQ056500 CIT-HSP-2
28	15.8	79.0	449	12 AZ958138	AZ958138 2M0225J18
29	15.8	79.0	450	9 AW185013	AW185013 se85e08.y
30	15.8	79.0	452	9 AI143727	AI143727 qf25407.x
31	15.8	79.0	453	12 AQ637264	AQ637264 RPCI-11-4
32	15.8	79.0	456	9 BE105642	BE105642 UI-R-BX0-
33	15.8	79.0	456	12 AZ485128	AZ485128 1M0312K14
34	15.8	79.0	458	9 AI275146	AI275146 ql70h09.x
35	15.8	79.0	459	9 BB862616	BB862616 BB862616
36	15.8	79.0	473	10 H08472	H08472 Y194a01.r1
37	15.8	79.0	476	10 W24943	W24943 2B66C07.r1
38	15.8	79.0	481	9 AI586294	AI586294 vr99d11.x
39	15.8	79.0	481	10 BG376470	BG376470 UI-R-CU0-
40	15.8	79.0	484	12 AQ461629	AQ461629 HS-5204.A
41	15.8	79.0	498	12 AZ215840	AZ215840 Sheared.D
42	15.8	79.0	506	12 AZ248018	AZ248018 RPCI-23-4
43	15.8	79.0	511	10 B1467618	B1467618 389271.MA
44	15.8	79.0	521	9 AI798758	AI798758 we92C09.x
45	15.8	79.0	536	10 BM391560	BM391560 UI-R-DY0-

#### ALIGNMENTS

#### RESULT 1

AG045201 657 bp DNA linear GSS 02-NOV-2001  
Pan troglodytes DNA, clone: PTB-024A19.F, genomic survey sequence.  
AG045201  
AG045201.1 GI:16582018  
GSS: GSS (genome survey sequence).  
Pan troglodytes male lymphoblast DNA, clone\_lib:PTB Chimpanzee Male  
BAC Library clone:PTB-024A19.F.  
Pan troglodytes  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

#### REFERENCE

1 (sites)  
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Totoki,Y., Watanabe,H. and Sakaki,Y.  
BAC end sequences of Library PTB  
Unpublished  
2 (bases 1 to 657)  
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Totoki,Y., Watanabe,H. and Sakaki,Y.  
Direct Submission

#### TITLE

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail:chimpesegsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,  
Tel:81-45-503-9111, Fax:81-45-503-9170)

#### COMMENT

Clones are derived from the chimpanzee BAC library PTB This BAC end  
was generated during the R&D process and may have higher chance of  
clone tracking errors.

#### PRIMERS

Sequencing: -21M13

#### LIBRARY

Vector : pKS145  
R.Site 1 : SacI  
R.Site 2 : SacI  
Location/Qualifiers  
1. .657  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="PTB-024A19.F"

#### FEATURES

source

```

/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
151 a 156 c 186 g 164 t
BASE COUNT
ORIGIN

Query Match 100.0%; Score 20; DB 12; Length 657;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccctttacacctaagc 20
|||||
Db 159 CTCCCTTTACACCTAAGGC 178
CTCCCTTTACACCTAAGGC 178

RESULT 2
AG126122
LOCUS
DEFINITION Pan troglodytes DNA, clone: PTB-136J09.K, genomic survey sequence.
ACCESSION AG126122
VERSION AG126122.1 GI:16655287
KEYWORDS GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male BAC Library clone:PTB-136J09.K.
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
1 (sites)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of Library PTB
Unpublished
2 (bases 1 to 688)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimbes@sc.riken.go.jp, URL: http://hqp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1..688
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-136J09.K"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
1 others
BASE COUNT 148 a 174 c 187 g 178 t
ORIGIN

Query Match 100.0%; Score 20; DB 12; Length 688;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccctttacacctaagc 20
|||||
Db 146 CTCCCTTTACACCTAAGGC 165
CTCCCTTTACACCTAAGGC 165

RESULT 3

```

```

BB621639
LOCUS
DEFINITION BB621639 RIKEN full-length enriched, 11 days embryo head Mus
musculus cDNA clone 6230412E16 5', mRNA sequence.
ACCESSION BB621639
VERSION BB621639.1 GI:16460292
KEYWORDS EST.
SOURCE house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 653)
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,
Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda,
M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,
Okazaki, Y., Okido, T., Saito, K., Sakai, C., Sakai, K., Sano, H., Sasaki,
D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toyota, F.,
Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
non-redundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamana, I., Aizawa
K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
Hayashizaki, Y.
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
Location/Qualifiers
1..653
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="6230412E16"
/clone_lib="RIKEN full-length enriched, 11 days embryo
head"
/sex="mixed"
/tissue_type="head"
/dev_stage="11 days embryo"
/lab_host="DH10B"
/Note="Site 1: Sali; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was

```



Email: smithemail.marc.usda.gov  
 Single pass sequencing. Bases called and alt\_trimmed with phred  
 v0.980904.e. Vector identified by cross\_match with the -minscore 18  
 and -minmatch 12 options.

## PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCCAGTCAGCAGC

Plate: 86 row: 1 column: 8

Seq primer: ATTTAGGTGACACTATAG.

## FEATURES

source

Location/Qualifiers

1..420

/organism="Sus scrofa"

/db\_xref="taxon:9823"

/clone\_lib="MARC 2Pig"

/tissue\_type="pooled"

/lab\_host="DH10B"

/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
 Library made from pooled tissue from testis, ovary,  
 endometrium, hypothalamus, pituitary, and placenta."

129 a 73 c 91 y 127 t

BASE COUNT

ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 420;

Best Local Similarity 90.0%; Pred. No. 4.6e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctccctttacacctaagc 20

||||||| ||| |||||

Db 143 CTCCTTTTGAACCTAAGGC 124

## RESULT 7

AZ289149/c

LOCUS

DEFINITION

RPIC1-23-59B12.TVB RPIC1-23 Mus musculus genomic clone RPIC1-23-59B12,

DNA sequence.

ACCESSION AZ289149

VERSION 1

KEYWORDS

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheraia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 606)

Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret

, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.

and Fraser, C.W.

Mouse BAC End Sequences from Library RPIC1-23

Unpublished (1999)

Other\_GSSs: RPIC1-23-59B12.TJB

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhaoetigr.org

Clones are derived from the mouse BAC library RPIC1-23. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)

or from Resea ch Genetics (info@resgen.com). BAC end page:

http://www.tigr.org/tdb/bac/mouse/bac\_end\_intro.html

Plate: 59 row: B column: 12

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1..606

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPIC1-23-59B12"

## FEATURES

source

/clone\_lib="RPIC1-23"

/sex="Female"

/lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site\_1:  
 EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or  
 brain genomic DNA was isolated and partially digested  
 with a combination of EcoRI and EcoRI Methylase. Size  
 selected DNA was cloned into the pBACe3.6 vector at the  
 EcoRI sites. The ligation products were transformed into  
 DH10B electrocompetent cells (BRL Life Technologies)."

129 a 153 c 135 g 189 t

BASE COUNT

ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 606;

Best Local Similarity 90.0%; Pred. No. 4.9e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctccctttacacctaagc 20

||||||| ||| |||||

Db 328 CTCCTTTTAAACCAAGGC 309

## RESULT 8

AZ787637

LOCUS

DEFINITION

2M0034N09F Mouse 10kb plasmid UUC1M library Mus musculus genomic

clone UUCG2M0034N09 F, DNA sequence.

ACCESSION AZ787637

VERSION AZ787637.1

KEYWORDS

SOURCE

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheraia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 632)

Dunn, H., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly

, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.

and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0034 row: N column: 09

Seq primer: CGTTGTAACGACGGCCAGT

Class: plasmid ends

High quality sequence stop: 632.

Location/Qualifiers

1..632

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUCG2M0034N09"

/clone\_lib="Mouse 10kb plasmid UUC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"

/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The  
 adapted DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pWD42 (g11473211419b1AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adapted mouse DNA was annealed to  
 adapted vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

BASE COUNT 190 a 129 c 175 g 138 t  
 ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 632;  
 Best Local Similarity 90.0%; Pred. No. 4.9e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctccctttacaacctaagc 20  
 ||||| ||||| ||||| |||||  
 Db 515 CTCCCTTTAAACCAAGGC 534

RESULT 9  
 BE271342 891 bp mRNA linear EST 26-OCT-2000  
 LOCUS 601140504F1 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3049753 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BE271342 GI:9145026  
 VERSION BE271342  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 891)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 JOURNAL  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: DCTD/DTF  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov  
 Plate: L1CM94 row: g column: 02  
 High quality sequence start: 2  
 High quality sequence stop: 199.  
 High quality sequence stop: 199.  
 Location/Qualifiers

FEATURES  
 source  
 1..891  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3049753"  
 /clone\_lib="NIH\_MGC\_9"  
 /tissue\_type="adenocarcinoma cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: ovary; Vector: POTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dr priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACGAG(C). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 318 a 292 c 103 g 178 t  
 ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 891;  
 Best Local Similarity 90.0%; Pred. No. 5.2e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctccctttacaacctaagc 20  
 ||||| ||||| ||||| |||||  
 Db 477 CTACCTTTACCACCTAAGGC 496

RESULT 10  
 BE917358/c 904 bp mRNA linear EST 29-SEP-2000  
 LOCUS 601566260F1 NCI\_CGAP\_Mam1 Mus musculus cDNA clone IMAGE:3966374 5',  
 DEFINITION mRNA sequence.

ACCESSION BE917358  
 VERSION BE917358  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 904)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 JOURNAL  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1AM939 row: g column: 15  
 High quality sequence stop: 724.  
 High quality sequence stop: 724.  
 Location/Qualifiers

FEATURES  
 source  
 1..904  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3966374"  
 /clone\_lib="NCI\_CGAP\_Mam1"  
 /tissue\_type="tumor, biopsy sample"  
 /dev\_stage="10 months, virgin"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
 Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Library constructed by Life Technologies. Investigator  
 providing samples: Gilbert Smith, NIH"  
 BASE COUNT 212 a 232 c 241 g 219 t  
 ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 904;  
 Best Local Similarity 90.0%; Pred. No. 5.2e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctccctttacaacctaagc 20  
 ||||| ||||| ||||| |||||  
 Db 875 CTCCCTTTACTACCTAAGTC 856

RESULT 11  
 BG259805 975 bp mRNA linear EST 13-FEB-2001  
 LOCUS 602371951F1 NIH\_MGC\_93 Homo sapiens cDNA clone IMAGE:4479756 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG259805  
 VERSION BG259805.1 GI:12769621  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.





/db\_xref="taxon:3712"  
/clone="BOGWH69"  
/clone\_lib="BOGW"  
/note="Vector: pHOS1; Site\_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"  
BASE COUNT 200 a 132 c 176 g 220 t  
ORIGIN

Query Match 82.0%; Score 16.4; DB 12; Length 728;  
Best Local Similarity 94.4%; Pred. No. 8e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 3 ccccttacaacctaaaggc 20  
||| ||||| ||||| |||||  
Db 222 CCATTACACCTAAGGC 205

RESULT 14  
BG965724/c  
LOCUS 734 bp mRNA linear EST 12-JUN-2001  
DEFINITION 602830580F1 NCI\_CGAP\_Co24 Mus musculus cDNA clone IMAGE:4985347 5', mRNA sequence.  
ACCESSION BG965724  
VERSION BG965724.1 GI:14353361  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 734)  
REFERENCE NTH-MGC http://mgc.nci.nih.gov/  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Plate: LLAM1092 row: p column: 20  
High quality sequence stop: 734.  
Location/Qualifiers  
1..734  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4985347"  
/clone\_lib="NCI\_CGAP\_Co24"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: Colon; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP Library."  
BASE COUNT 191 a 175 c 202 g 166 t  
ORIGIN

Query Match 82.0%; Score 16.4; DB 10; Length 734;  
Best Local Similarity 94.4%; Pred. No. 8e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ctcctttacaacctaaag 18  
||| ||||| ||||| |||||  
Db 658 CTCCTTTCCACCTAAG 641

RESULT 15  
BF341631  
LOCUS 801 bp mRNA linear EST 22-NOV-2000

DEFINITION 602016062F1 NCI\_CGAP\_Brn64 Homo sapiens cDNA clone IMAGE:4151942 5', mRNA sequence.  
ACCESSION BF341631  
VERSION BF341631.1 GI:11288127  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 801)  
REFERENCE NIH-MGC http://mgc.nci.nih.gov/  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: David N. Louis, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Plate: LLAM9417 row: k column: 15  
High quality sequence start: 3  
High quality sequence stop: 400.  
Location/Qualifiers  
1..801  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4151942"  
/clone\_lib="NCI\_CGAP\_Brn64"  
/tissue\_type="gliblastoma with EGFR amplification"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: Brain; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.57 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP Library."  
BASE COUNT 209 a 245 c 143 g 204 t  
ORIGIN

Query Match 82.0%; Score 16.4; DB 10; Length 801;  
Best Local Similarity 94.4%; Pred. No. 8.1e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ctcctttacaacctaaag 18  
||| ||||| ||||| |||||  
Db 416 CTCCTTTACACCTAAG 433

Search completed: May 18, 2002, 17:22:02  
Job time: 14673 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:45:53 ; Search time 1024.22 Seconds  
(without alignments)  
50.289 Million cell updates/sec

Title: US-09-623-329-5

Perfect score: 30

Sequence: 1 cgtctcccttggatggccctggacc 30

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

1: /SID55/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID55/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID55/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID55/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID55/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID55/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID55/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID55/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID55/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID55/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID55/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID55/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID55/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID55/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID55/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID55/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	100.0	30	AAZ11655	Epstein Barr Nucle
2	30	100.0	624	AAT15397	PCR-generated eEBN
3	30	100.0	1092	AAZ03478	Epstein-Barr nucle
4	30	100.0	1212	AAT15396	EBV nuclear antige
5	30	100.0	1748	AAZ03475	Epstein-Barr virus
6	30	100.0	1925	AAZ09024	Epstein Barr virus
7	30	100.0	1926	AAA50254	Epstein Barr virus
8	30	100.0	1926	AAF82902	EBV tethering prot
9	30	100.0	2580	AAA75454	Nucleotide sequenc

c	10	30	100.0	5452	20	AAZ09023	Anti-sense strand
c	11	30	100.0	8575	14	AAZ040324	Sequence of PHEB03
c	12	30	100.0	8705	20	AAZ23778	Vector pShuttle DN
c	13	30	100.0	9600	19	AAZ21683	Vector plasmid PCM
c	14	30	100.0	9725	21	AAZ53873	Expression vector
c	15	30	100.0	9732	21	AAZ53879	Expression vector
c	16	30	100.0	9738	21	AAZ53874	Expression vector
c	17	30	100.0	9873	21	AAZ53875	Expression vector
c	18	30	100.0	10054	21	AAZ53876	Expression vector
c	19	30	100.0	10380	20	AAZ22248	Nucleotide sequenc
c	20	30	100.0	10596	14	AAZ051731	Plasmid pcISEBON f
c	21	30	100.0	10596	17	AAT40348	Plasmid pcISEBON f
c	22	30	100.0	10596	20	AAZ10650	Nucleotide sequenc
c	23	30	100.0	11265	19	AAV59501	Plasmid pREP1:CTL
c	24	30	100.0	16080	21	AAZ59553	DNA clone pCEK C1.
c	25	20.4	68.0	2613	22	AAH17468	Human cDNA sequenc
c	26	20	66.7	20	14	AAO52595	EBV target sequenc
c	27	19.4	64.7	2001	23	AAZ591879	DNA encoding novel
c	28	19.2	64.0	7284	22	ABA15817	Human nervous syst
c	29	19	63.3	614	22	AAK73435	Human immune/haema
c	30	19	63.3	737	22	AAK58707	Human immune/haema
c	31	19	63.3	4182	22	AAH26524	Human proten/oligo
c	32	19	63.3	26555	22	AAK68372	Human immune/haema
c	33	19	63.3	26555	22	AAK68605	Human immune/haema
c	34	19	63.3	26555	22	AAI62833	Human genomic DNA
c	35	18.8	62.7	275	22	ABA14827	Human nervous syst
c	36	18.8	62.7	1015	21	AAA98825	Human proliferatio
c	37	18.8	62.7	1015	21	AAA981	Human proliferatio
c	38	18.8	62.7	1158	15	AAQ7137	Human developmenta
c	39	18.8	62.7	1317	22	AAH33430	Human colon cancer
c	40	18.8	62.7	2429	22	AAH17367	Human cDNA sequenc
c	41	18.8	62.7	2550	15	AAQ71335	Mature human devel
c	42	18.8	62.7	3074	21	AAQ76994	Human ORFX ORF2549
c	43	18.8	62.7	3611	16	AAQ94421	Human Rse rPTK cDN
c	44	18.8	62.7	3611	21	AAA99915	cDNA encoding huma
c	45	18.8	62.7	3842	23	ABL15059	Drosophila melanog

#### ALIGNMENTS

RESULT 1

AAZ11655

ID AAZ11655 standard; DNA; 30 BP.

XX AAZ11655;

AC AAZ11655;

XX 19-NOV-1999 (first entry)

DT 19-NOV-1999 (first entry)

DE Epstein Barr Nuclear Antigen 1 (EBNA-1) specific probe.

XX Epstein Barr Virus; EBV infection; viral; gene transcription; EBNA-1;

XX Epstein Barr early RNA; Epstein Barr nuclear antigen 1; EBNA-1; LMP-1;

KW latent membrane protein; LMP-2; vIL10; BCRF-1; BARF1; BDLF2; NASBA;

KW EBV-associated malignancy; probe; ss.

XX Synthetic.

OS Epstein-barr virus.

XX WO9945155-A2.

PN 10-SEP-1999.

PD 10-SEP-1999.

XX 01-MAR-1999; 99WO-EF01392.

XX 04-MAR-1998; 98EP-0200655.

PR 14-DEC-1998; 98EP-0204231.

XX (ALKU ) AKZO NOBEL NV.

PI Vervooort MBHJ, Van Den Brule AJC, Middelidorp JM;

XX WPI; 1999-551051/46.

XX Identifying Epstein Barr Virus infection -  
 PT Claim 14; Page 22; 50pp; English.  
 PS  
 CC The invention provides methods for identifying an Epstein Barr Virus  
 CC (EBV) infection, that comprises determining viral gene transcription  
 CC patterns by amplification of specific RNA sequences. The binding sites  
 CC of the oligos suitable for amplification are located in the following  
 CC genes: Epstein Barr early RNA (EBER-1), Epstein Barr nuclear antigen 1  
 CC (EBNA-1), latent membrane protein 1 (LMP-1), LMP-2, and VIL10 (BCRF-1),  
 CC BARF1 and BDLF2. The method comprises (a) amplifying a target sequence  
 CC within one or more RNA(s) transcribed from above gene sequences and the  
 CC pattern and the amplified products, determining the transcription  
 CC pattern and identifying the corresponding EBV-associated malignancy. The  
 CC RNA is amplified using a transcription based amplification technique such  
 CC as NASBA. The invention is used to diagnose malignant and non-malignant  
 CC EBV-associated diseases. The present sequence represents a probe specific  
 CC for EBNA-1 RNA comprising a detectable label.  
 XX Sequence 30 BP; 3 A; 13 C; 7 G; 7 T; 0 other;  
 SQ

Query Match 100.0%; Score 30; DH 20; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 0.0012;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgtctccctttggaatgccccctggacc 30  
 |||||  
 DB 1 cgtctccctttggaatgccccctggacc 30

RESULT 2  
 AAT15397  
 ID AAT15397 standard; DNA; 624 BP.

XX AAT15397;  
 XX 19-JUL-1996 (first entry)  
 DT PCR-generated eEBNA1 protein gene.  
 DE  
 DE EBV; nuclear antigen protein 1; EBNA1; immunoassay; ds.  
 XX Epstein-barr virus.  
 OS

XX Key Location/Qualifiers  
 FH primer\_bind complement (1...24)  
 FT /\*tag= a  
 FT /note= "primer for eEBNA1"  
 FT 584..624  
 FT /\*tag= b  
 FT /note= "primer for eEBNA1"

XX WO9602563-A1.

XX 01-FEB-1996.

XX 13-JUL-1995; 95WO-US08700.

XX 13-JUL-1994; 94US-0275614.

XX (CORR ) CORNELL RES FOUND INC.

XX O'Donnell ME;

XX P-PSDB; AAR88588.

XX Recovery of Epstein Barr Virus nuclear antigen protein 1 (EBNA1) -

PT useful in immunoassay system for detecting Epstein Barr Virus in  
 PT serum samples  
 XX

PS Example 22; Page 51; 82pp; English.

XX A DNA fragment (AAT15397) was generated by PCR amplification of  
 CC the Epstein-Barr virus (EBV) nuclear antigen protein 1 (EBNA1)  
 CC gene (see also AAT15396) in plasmid p291 and NdeI/BamHI digestion of  
 CC the PCR product. It was ligated into pET3c and the resulting  
 CC plasmid was used to transform Escherichia coli strain BL21  
 CC (DE3)plyss cells. The recombinant EBNA1, designated eEBNA1  
 CC (AAR88588), can be recovered from the nuclei of host cells for use  
 CC EBV immunoassays.

XX Sequence 624 BP; 148 A; 132 C; 195 G; 149 T; 0 other;

Query Match 100.0%; Score 30; DB 17; Length 624;  
 Best Local Similarity 100.0%; Pred. No. 0.0017;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgtctccctttggaatgccccctggacc 30  
 |||||  
 DB 289 cgtctccctttggaatgccccctggacc 318

RESULT 3  
 AAX03478  
 ID AAX03478 standard; DNA; 1092 BP.

XX AAX03478;  
 AC

XX 26-APR-1999 (first entry)  
 DT

XX Epstein-Barr nuclear antigen 1 (EBNA1) coding region.

DE EBV; replicon; episomal replication; Orip; EBNA1;

KW Epstein-Barr nuclear antigen 1; gene therapy; vector; ss.

XX Epstein-Barr virus.

XX WO9857658-A1.

XX 23-DEC-1998.

XX 14-APR-1998; 98WO-US07641.

XX 19-JUN-1997; 97US-0050206.

XX (GENO ) GEN HOSPITAL CORP.

XX Seed B;

XX WPI; 1999-080933/07.

XX Epstein-Barr virus replicons which support episomal replication -  
 PT comprise an Orip sequence and an EBNA1 sequence operably linked to a  
 PT promoter

XX Claim 4; Page -; 12pp; English.

XX This is the Epstein-Barr nuclear antigen 1 (EBNA1) coding region.  
 CC A novel, claimed nucleic acid sequence which supports episomal  
 CC replication in a mammalian cell comprises an Epstein-Barr virus  
 CC (EBV) Orip sequence (see AAX03477) and an EBNA1 sequence operably  
 CC linked to a promoter, the whole sequence having a length of less  
 CC than 3 kb, and preferably less than 1.8 kb (see AAX03475). The  
 CC nucleic acid sequence preferably further comprises a bidirectional  
 CC polyadenylation consensus (see AAX03476), and is designed to be  
 CC inserted upstream of such a polyadenylation sequence in an  
 CC appropriate vector. The claimed nucleic acid sequence acts as a  
 CC compact EBV replicon. Plasmid vectors based on this sequence  
 CC replicate as episomes in the nucleus of transfected mammalian  
 CC cells. Compact EBV replicons find use in gene therapy vectors,  
 CC e.g. in gene delivery vehicles such as expression vectors.  
 CC (NB. the sequence of EBNA1 was created with respect to Claim 4 and

CC the sequence provided in Fig 1 of the specification).  
 XX  
 SQ Sequence 1092 BP; 266 A; 236 C; 393 G; 197 T; 0 other;

Query Match 100.0%; Score 30; DB 20; Length 1092;  
 Best Local Similarity 100.0%; Pred. No. 0.0018;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 cgtctcccttgggaatggccctggacc 30  
 Db 781 cgtctcccttgggaatggccctggacc 810

## RESULT 4

AAT15396  
 ID AAT15396 standard; DNA; 1212 BP.  
 XX  
 AC AAT15396;  
 XX  
 DT 19-JUL-1996 (first entry)  
 XX  
 DE EBV nuclear antigen protein 1 gene.  
 XX  
 KW EBV; nuclear antigen protein 1; EBNA1; immunoassay; baculovirus;  
 KW ds.  
 XX  
 OS Epstein-barr virus.  
 XX  
 PN WO9602563-A1.  
 XX  
 PD 01-FEB-1996.  
 XX  
 PF 13-JUL-1995; 95WO-US08700.  
 XX  
 PR 13-JUL-1994; 94US-0275614.  
 XX  
 PA (CORR ) CORNELL RES FOUND INC.  
 XX  
 PI O'Donnell ME;  
 XX  
 DR WPI; 1996-105853/11.  
 DR P-FSDB; AAR88587.

PT Recovery of Epstein Barr Virus nuclear antigen protein 1 (EBNA1) -  
 PT useful in immunoassay system for detecting Epstein Barr Virus in  
 PT serum samples

PS Claim 12; Page 49; 82pp; English.

XX An isolated DNA (AAT15396) codes for the Epstein-Barr virus (EBV)  
 CC nuclear antigen protein 1 (EBNA1) (AAR88587), the viral encoded  
 CC protein which binds the latent phase origin (oriP) of EBV.  
 CC A recombinant DNA expression system comprises an expression  
 CC vector carrying an insert of heterologous DNA encoding EBNA-1.  
 CC The host cells may be e.g. Escherichia coli (see also AAT15397) or  
 CC Sf-9 insect cells (baculovirus expression system). The EBNA1 is  
 CC recovered from the host cell nucleus and used in EBV immunoassays.

XX  
 SQ Sequence 1212 BP; 307 A; 266 C; 429 G; 210 T; 0 other;

Query Match 100.0%; Score 30; DB 17; Length 1212;  
 Best Local Similarity 100.0%; Pred. No. 0.0018;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 cgtctcccttgggaatggccctggacc 30  
 Db 898 cgtctcccttgggaatggccctggacc 927

## RESULT 5

AAX03475

ID AAX03475 standard; DNA; 1748 BP.  
 XX  
 AC AAX03475;  
 XX  
 DT 26-APR-1999 (first entry)  
 XX  
 DE Epstein-Barr virus compact replicon.  
 XX  
 KW EBV; replicon; episomal replication; OriP; EBNA1;  
 KW Epstein-Barr nuclear antigen 1; gene therapy; vector; ss.  
 OS Chimeric - Epstein-Barr virus.  
 OS Chimeric - Herpes simplex virus type 1.  
 XX  
 PH Key Location/Qualifiers  
 FT misc\_feature 1..495  
 XX /\*tag= a  
 FT /note= "OriP, specifically claimed in Claim 3"  
 FT 496..616  
 FT /note= "herpes simplex virus 1 thymidine kinase  
 FT gene modified promoter"  
 FT CDS 627..1721  
 FT /\*tag= b  
 FT /product= Epstein-Barr nuclear antigen 1  
 FT /note= "EBNA1, specifically claimed in Claim 4"  
 XX  
 PN WO9857658-A1.  
 XX  
 PD 23-DEC-1998.  
 XX  
 PF 14-APR-1998; 98WO-US07641.  
 XX  
 PR 19-JUN-1997; 97US-0050206.  
 XX  
 PA (GEHO ) GEN HOSPITAL CORP.  
 XX  
 PI Seed B;  
 XX  
 DR WPI; 1999-080333/07.  
 XX  
 PT Epstein-Barr virus replicons which support episomal replication -  
 PT comprise an OriP sequence and an EBNA1 sequence operably linked to a  
 PT promoter  
 XX  
 PS Disclosure; Fig 1; 12pp; English.  
 XX  
 CC This 1748 bp fragment acts as a compact Epstein-Barr virus (EBV)  
 CC replicon. Extensive deletion and mutagenesis studies showed that  
 CC it is possible to embed the cis- and trans-acting functions  
 CC necessary for EBV episomal replication into a fragment of less than  
 CC 2 kb. The 1748 bp fragment contains all sequences needed for  
 CC efficient expression of the Epstein-Barr nuclear antigen 1 (EBNA-1)  
 CC protein, with the exception of a polyA consensus sequence. It is  
 CC designed to be inserted upstream of a bidirectional polyA sequence  
 CC (see also AAX03476) in an appropriate vector. Plasmid vectors based  
 CC on this sequence replicate as episomes in the nucleus of transfected  
 CC mammalian cells. Compact EBV replicons find use in gene therapy  
 CC vectors, e.g. in gene delivery vehicles such as expression vectors.  
 XX  
 SQ Sequence 1748 BP; 453 A; 382 C; 546 G; 367 T; 0 other;

Query Match 100.0%; Score 30; DB 20; Length 1748;  
 Best Local Similarity 100.0%; Pred. No. 0.0019;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 cgtctcccttgggaatggccctggacc 30  
 Db 1407 cgtctcccttgggaatggccctggacc 1436

## RESULT 6

AAX90924

AAAX90924 standard; DNA; 1925 BP.  
AAAX90924;  
17-JAN-2000 (first entry)  
Epstein Barr Virus Nuclear Antigen 1 (EBNA 1) DNA.  
Epstein Barr Virus Nuclear Antigen 1 DNA; plasmid pCMVEBNA; EBNA 1;  
episome; transfection; origin of replication; EBV orip; receptor;  
eukaryotic host cell; recombinant cell line; ion channel; gene therapy;  
multiple gene expression; transporter protein; transcription factor;  
adhesion molecule; antisense therapy; gene amplification;  
cell immortalisation; ds.  
Epstein-barr virus.  
Key Location/Qualifiers  
CDS 1..1925  
/tag= a "EBNA 1"  
/product= "EBNA 1"  
/trans\_except= (pos:799..800, aa:Gly)  
/note= "The sequence is described throughout the  
specification as being 1926 nucleotides long, but a  
sequence of only 1925 bp has been given in figure 2"  
WO9947647-A1.  
23-SEP-1999. 99WO-US03307.  
12-FEB-1999; 98US-0040961.  
18-MAR-1998; 98US-0130114.  
06-AUG-1998; 98US-0130114.  
(PHAR-) PHARMACOEPIA INC.  
Dama BB, Horlick RA, Robbins AK;  
WPI: 1999-610610/52.  
P-PSDB; AAY28843.  
New method for expressing genes from recombinant eukaryotic cells,  
useful for gene therapy -  
Claim 24; Fig 2; 86pp; English.  
The present sequence is a DNA encoding Epstein Barr Virus Nuclear  
Antigen 1 (EBNA 1), which is obtained from commercially available  
plasmid pCMVEBNA. EBNA 1 protein is used to stably maintain episomes  
containing EBV origin of replication (orip) and a gene encoding  
protein or RNA of interest. Eukaryotic host cells expressing EBNA 1  
protein are transfected with these episomes to produce recombinant  
cell lines expressing multiple genes of interest. This provides a  
rapid and reliable method of stably expressing multiple genes in  
transfected cells. The episomes are useful in the transfection of genes  
encoding receptors, transporter proteins, ion channels, adhesion  
molecules and transcription factors. The episomes carrying desired genes  
can also be used to transfect cells in gene therapy, antisense therapy,  
for gene amplification, cell immortalisation, etc.  
Sequence 1925 BP; 486 A; 352 C; 872 G; 215 T; 0 other;  
Query Match 100.0%; Score 30; DB 20; Length 1925;  
Best Local Similarity 100.0%; Pred. No. 0.0019;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 cgtctccctttggaatggcccttgacc 30  
Db 1611 cgtctccctttggaatggcccttgacc 1640

RESULT 7  
AAAS0254  
ID AAA50254 standard; DNA; 1926 BP.  
XX  
AC AAA50254;  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE Epstein Barr virus nuclear antigen 1 protein (EBNA1) DNA.  
XX  
KW EBV; nuclear antigen 1; EBNA1; episome; transfection; selection;  
KW gene therapy; ds.  
XX  
OS Epstein-barr virus.  
XX  
PN WO200047778-A1.  
XX  
PD 17-AUG-2000.  
XX  
PF 11-FEB-2000; 2000WO-US03547.  
XX  
PR 11-FEB-1999; 99US-0249585.  
XX  
PA (PHAR-) PHARMACOEPIA INC.  
XX  
PI Horlick RA, Chelsky D;  
XX  
DR WPI: 2000-515062/46.  
DR P-PSDB; AAY95856.  
XX  
PT Stably transfecting eukaryotic cells with at least one episome for the  
PT production of a desired protein in vitro and for gene therapy -  
PS Disclosure; Fig 2; 53pp; English.  
XX  
CC The present sequence is that of DNA encoding the Epstein-Barr virus  
CC (EBV) nuclear antigen protein 1 (EBNA1, see AAY95856). EBNA1 is  
CC utilised in a novel method for obtaining a eukaryotic cell that is  
CC stably transfected with at least one episome. This method involves  
CC transfecting a eukaryotic cell with: (1) a first episome comprising  
CC an EBV origin of replication (orip, see AAA50253), a gene encoding a  
CC first protein whose expression results in cell death and a  
CC selectable marker for eukaryotic cells; and (2) a second episome  
CC comprising an EBV orip and a gene encoding a second protein, where  
CC expression of the second protein prohibits the occurrence of cell  
CC death resulting from expression of the first protein to produce  
CC doubly transfected cells which also express an antigen that  
CC promotes retention of the episomes by the cells. The doubly  
CC transfected cells are maintained under conditions in which the  
CC first and second proteins and the selectable marker are expressed,  
CC and the selective pressure specified by the marker is maintained.  
CC Under these conditions, only cells containing both episomes live.  
CC Preferably, EBNA1 is expressed from 1 of the episomes, and the  
CC protein of interest from the other episome. Either or both  
CC epitopes may further comprise a nucleic acid sequence encoding a  
CC protein desired to be expressed in the cell (e.g. a therapeutic  
CC protein), a nucleic acid encoding an RNA that is not intended to  
CC be translated (e.g. a therapeutic RNA), or a DNA sequence used as  
CC a tag for the cells. The method is applicable to cell culture or  
CC intact organisms, for gene therapy. It allows the rapid  
CC establishment of eukaryotic cells that stably and reliably express  
CC a gene of interest, using a novel method of selection, and  
CC maintenance of that selection without the need for exogenous  
CC selection factors, such as antibiotics.  
XX  
SQ Sequence 1926 BP; 487 A; 352 C; 872 G; 215 T; 0 other;  
XX  
Query Match 100.0%; Score 30; DB 21; Length 1926;  
Best Local Similarity 100.0%; Pred. No. 0.0019;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 cgtctccctttggaatggcccttgacc 30

Db 1612 cgtctcccttgggaatggccctggacc 1641  
|||||

## RESULT 8

AAF82902

ID AAF82902 standard; DNA; 1926 BP.

XX AAF82902;

DT 29-JUN-2001 (first entry)

XX EBV tethering protein EBNA1 encoding DNA.

DE Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;  
KW Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;  
KW EBV; latency-associated nuclear antigen; LANA; EBNA1; ds.

XX Epstein-barr virus.

XX Key Location/Qualifiers

FH 1..1926

FT CDS /\*tag= a

PN W0200125484-A2.

XX 12-APR-2001.

XX 29-SEP-2000; 2000WO-US26908.

XX 01-OCT-1999; 99US-0410399.

XX (UNM1 ) UNIV MICHIGAN.

PA Robertson ES, Cotter MA;

PI WPI; 2001-281736/29.

DR P-PSDB; AAB62332.

XX A composition for use in gene therapy comprises an expression vector  
PT that includes a nucleic acid sequence encoding a nucleic acid binding  
PT protein -

XX Disclosure; Fig 9C; 60pp; English.

XX The invention provides a composition comprising nucleic acid, histone H1  
CC protein and expression vector operatively encoding a protein suitable  
CC for tethering the nucleic acid to the histone H1 protein, where the  
CC tethering protein is LANA. The composition is useful in aiding the  
CC retention of the viral DNA in the host cell. The viral vector encodes a  
CC protein suitable for tethering DNA to Histone H1. Methods for screening  
CC for compounds which are agonistic or antagonistic for the tethering of  
CC viral proteins to histone H1 and DNA binding sites are useful for  
CC developing the method of viral transfer. The composition has applications  
CC to gene therapy, including the treatment of multiple sclerosis,  
CC Parkinson's disease, Huntington disease and diabetes. The present  
CC sequence represents the nucleotide sequence of the Epstein-barr virus  
CC (EBV) tethering protein ENNA1.

XX Sequence 1926 BP; 487 A; 352 C; 872 G; 215 T; 0 other;

## Query Match

Best Local Similarity 100.0%; Score 30; DB 22; Length 1926;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgtctcccttgggaatggccctggacc 30

|||||

Db 1612 cgtctcccttgggaatggccctggacc 1641

## RESULT 9

AAA75454

ID AAA75454 standard; DNA; 2580 BP.

XX AAA75454;

DT 15-JAN-2001 (first entry)

XX Nucleotide sequence of the Epstein Barr nuclear antigen.

DE Origin of replication; protein-protein interaction; replication;  
KW two-hybrid system; nuclear antigen; ss.

XX Epstein-barr virus.

XX US6114111-A.

XX 05-SEP-2000.

XX 30-MAR-1998; 98US 0050863.

XX 30-MAR-1998; 98US-0050863.

XX (RIGE-) RIGEL PHARM INC.

XX Luo Y, Payan D, Huang B;

XX WPI; 2000-59,546/56.

XX Composition for detecting protein-protein interactions in a mammalian  
PT two hybrid system comprises bait and test vector which consist of  
PT selection gene, vector viral origin of replication and fusion gene -  
XX Disclosure; Column 17-20; 18pp; English.

XX The present sequence represents the Epstein Barr nuclear antigen.  
CC It is used to produce bait vectors of the invention. The specification  
CC describes a compositions and methods for a genetic system of detecting  
CC protein-protein interactions in a mammalian host cell. The system  
CC comprises bait and test, both containing selection genes, and viral  
CC origin of replications which require bound viral replication proteins  
CC to effect replication. The compositions is useful for detecting an  
CC interaction between a bait protein and a test protein. It is useful in  
CC a mammalian two-hybrid system for detecting protein-protein interactions  
CC in a mammalian host cell.

XX Sequence 2580 BP; 632 A; 512 C; 1054 G; 382 T; 0 other;

## Query Match

Best Local Similarity 100.0%; Score 30; DB 21; Length 2580;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgtctcccttgggaatggccctggacc 30

|||||

Db 1995 cgtctcccttgggaatggccctggacc 2024

## RESULT 10

AAAX90923/C

ID AAAX90923 standard; DNA; 5452 BP.

XX AAAX90923;

DT 17-JAN-2000 (first entry)

XX Anti-sense strand of pCMVEBNA plasmid.

DE Anti-sense strand; plasmid pCMVEBNA; EBNA 1; episome; transfection;  
KW Epstein Barr Virus Nuclear Antigen 1; origin of replication;  
KW EBV orip; eucaryotic host cell; recombinant cell line; ion channel;  
KW multiple gene expression; receptor; transporter protein; gene therapy;  
KW transcription factor; adhesion molecule; antisense therapy;  
KW gene amplification; cell immortalisation; ss.

```

OS Epstein-barr virus.
OS Cytomegalovirus.
XX Synthetic.
FH Key Location/Qualifiers
FT CDS complement (3032..4957)
FT /tag= a
FT /product= "EBNA 1 protein"
FT /note= "Epstein Barr Virus Nuclear Antigen 1"
XX
XX WO947647-A1.
XX
XX 23-SEP-1999.
XX
XX 12-FEB-1999; 99WO-US03307.
XX
XX 18-MAR-1998; 98US-0040961.
XX 06-AUG-1998; 98US-0130114.
XX (PHAR-) PHARMACOPEIA INC.
XX
XX Damaj BB, Horlick RA, Robbins AK;
XX WPI; 1999-610610/52.
XX
XX New method for expressing genes from recombinant eukaryotic cells,
XX useful for gene therapy -
XX
XX Example 1; Fig 1; 86pp; English.
XX
XX The present sequence is an anti-sense strand of commercially available
XX plasmid pCMVEBNA from which Epstein Barr Virus Nuclear Antigen 1 (EBNA
XX 1) encoding DNA is obtained. EBNA 1 protein is used to stably maintain
XX episomes containing EBV origin of replication (oriP) and a gene encoding
XX protein or RNA of interest. Eucaryotic host cells expressing EBNA 1
XX protein are transfected with these episomes to produce recombinant
XX cell lines expressing multiple genes of interest. This provides a
XX rapid and reliable method of stably expressing multiple genes in
XX transfected cells. The episomes are useful in the transfection of genes
XX encoding receptors, transporter proteins, ion channels, adhesion
XX molecules and transcription factors. The episomes carrying desired genes
XX can also be used to transfect cells in gene therapy, antisense therapy,
XX for gene amplification, cell immortalisation, etc.
XX
XX Sequence 5452 BP; 1108 A; 1736 C; 1246 G; 1362 T; 0 other;
XX
XX
XX Query Match 100.0%; Score 30; DR 20; Length 5452;
XX Best Local Similarity 100.0%; Pred No. 0.0021;
XX Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 cgtctcccttgggaatggcccttgagacc 30
Db |||||
Db 810 CGTCTCCCTTGGGAATGGCCCTGGACCC 781

RESULT 11
AAQ40324
XX AAQ40324 standard; cDNA; 8575 BP.
XX
XX AC AAQ40324;
XX
XX DT 09-AUG-1993 (first entry)
XX
XX DE Sequence of pHEB030.
XX
XX KW Neurotrophic factor; growth promoting activity; GPA; ss.
XX
XX OS Gallus domesticus.
XX
XX FH Key Location/Qualifiers
FT misc_feature 781..3061
FT /tag= a

```

```

FT repeat_region /label= orip
FT 882..1474
FT /tag= b
FT misc_feature 2587..2665
FT /tag= c
FT /label= dyad region
FT 3062
FT /tag= d
FT /note= "HSV TK Term 3' end"
FT 3159..3737
FT /tag= e
FT /label= CMV
FT 3738..2901
FT /tag= f
FT 2902..3988
FT /tag= g
FT /label= SP6
FT 3989..3993
FT /tag= h
FT /note= "Sp6 KNA start"
FT 4510..4646
FT /tag= i
FT /label= SV40 early poly A
FT /note= "g"
FT 4647..4691
FT /tag= j
FT /label= SV40 origin
FT 4692
FT /tag= k
FT /label= HSV TK Term 5' end
FT complement (5185..6207)
FT /tag= l
FT /label= HPG gene from pLG89-- hygrosistance
FT complement (6234..6279)
FT /tag= m
FT /label= HSV TK
FT complement (6306..6359)
FT /tag= n
FT CAAT_signal complement (6360..6479)
FT /tag= o
FT CDS complement (6480..6803)
FT /tag= p
FT /label= PBR 322; tel-kpeptide
FT complement (7604..8448)
FT /tag= q
FT /label= beta lastanase
XX
XX WO9307270-A.
XX
XX 15-APR-1993.
XX
XX 29-SEP-1992; 92WO-US08258.
XX
XX 01-OCT-1991; 91US-0769622.
XX
XX (GETH ) GENENTECH INC.
XX (UYOR-) UNIV OREGON HEALTH SCI.
XX
XX Cachianes G, Eckenstein PP, Leung D, Nishi R;
XX WPI; 1993-134459/16.
XX
XX Nucleic acid encoding growth promoting activity - used for
XX developing prods. for diagnosis and treatment of neurological
XX diseases and disorders
XX
XX Disclosure; Fig 4; 74pp; English.
XX
XX Eyes from embryonic day 15 (E15) chicken embryos were dissected and
XX total RNA was isolated. The RNA was used to prepare cDNA which was
XX than ligated to Sfil digested lambda-HEBO vector. The E15 cDNA
XX library was screened using oligonucleotide probes based on the
XX partial amino acid sequences of GPA obtd. by microsequencing of 3

```



CC different peptide fragments of purified GPA. The oligonucleotides  
 CC are referred to as o-GPA-1, o-GPA-2 and o-GPA-3. Lambda CEL5 #19  
 CC hybridized with all three probes. The expression of GPA involves  
 CC the use of a plasmid expression vector (pHEB030) containing the orip  
 CC region from ERV. pHEB030 comprises the strong CMV promoter, a multiple  
 CC cloning region for insertion of foreign (exogenous) genes downstream  
 CC of the CMV promoter, the orip region of EBV for plasmid replication  
 CC in host cells expressing EBNA-1, a hygromycin resistance gene for  
 CC selection in eukaryotes, the origin of replication from pBR322 for  
 CC replication in prokaryotes, and an ampicillin resistance gene for  
 CC selection in prokaryotes. In the sequence "N" is used to  
 CC designate the nucleotides that comprise the arbitrary 350 base pair  
 CC CDNA insert in pHEB030.  
 XX  
 SQ Sequence 8575 BP; 1989 A; 2167 C; 2032 G; 2037 T; 350 other;

Query Match 100.0%; Score 30; DB 14; Length 8575;  
 Best Local Similarity 100.0%; Pred. No. 0.0023;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgtctcccttggatggccctggacc 30  
 ||||||||||||||||||||||||  
 Db 166 cgtctcccttggatggccctggacc 195

RESULT 12  
 AA223778/c  
 ID AA223778 standard; DNA: 8705 BP.  
 AC  
 XX AA223778;  
 DT 14-JAN-2000 (first entry)  
 XX  
 DE Vector pshuttle DNA.  
 XX  
 KW Antisense; DNA library; identification; multiple cloning site; MCS;  
 KW inhibition; ss.  
 XX Synthetic.

OS  
 XX WO9950457-A1.  
 PN  
 XX 07-OCT-1999.  
 PD  
 XX 28-MAR-1999; 99WO-US06742.  
 PF  
 XX 28-MAR-1998; 98US-0079792.  
 PR 06-NOV-1998; 98US-0107504.  
 XX (UTAH ) UNIV UTAH RES FOUND.

XX Ruffner DE, Pierce ML, Chen Z;  
 XX WPI; 1999-610866/52.

XX Production of antisense libraries, used for identifying antisense  
 XX agents and for identifying target sites for antisense-mediated  
 XX inhibition of a selected gene -

XX Claim 16; Page 43-50; 63pp; English.

XX This invention describes a novel method for generating an antisense  
 CC library targeted to a selected RNA transcript. The methods can be used  
 CC for identifying antisense agents and for identifying target sites for  
 CC antisense-mediated inhibition of a selected gene. The use of a direct  
 CC library for target site selection significantly simplifies the screening  
 CC process, since only very small libraries need be prepared and assayed.  
 CC This sequence represents the vector pShuttle which is used in the method  
 CC of the invention.

XX Sequence 8705 BP; 1875 A; 2575 C; 2078 G; 2177 T; 0 other;

Query Match 100.0%; Score 30; DB 20; Length 8705;  
 Best Local Similarity 100.0%; Pred. No. 0.0023;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgtctcccttggatggccctggacc 30  
 ||||||||||||||||||||||||  
 Db 6675 cgtctcccttggatggccctggacc 6646

RESULT 13  
 AA221683  
 ID AAV21683 standard; DNA: 9600 BP.  
 XX  
 XX AAV21683;  
 DT 17-AUG-1998 (first entry)  
 XX  
 DE Vector plasmid pCMVKmITR-EPI.  
 XX  
 KW Polynucleotide delivery; plasmid pCMVKmITR-EPI; vector;  
 KW gene therapy; vaccine; polyclonal agent; ss.

XX Chimeric - Epstein-Barr virus.  
 OS Chimeric - Adeno-associated virus.  
 OS Chimeric - Cytomegalovirus.  
 OS Chimeric - Bos taurus.

XX Key Location/Qualifiers  
 FH 14..2594  
 CDS /\*tag= a  
 FT /\*product= "EBV nuclear antigen A"  
 FT misc\_feature 2623..4559  
 FT /\*tag= b  
 FT /\*note= "EBV origin of replication"  
 FT repeat\_unit 4928..5104  
 FT /\*tag= c  
 FT /\*rpt\_type= INVERTED  
 FT /\*note= "AAV inverted terminal repeat"  
 FT repeat\_unit 7189..7355  
 FT /\*tag= d  
 FT /\*rpt\_type= INVERTED  
 FT /\*note= "AAV inverted terminal repeat"  
 FT promoter 5112..6734  
 FT /\*tag= e  
 FT /\*note= "CMV immediate-early enhancer/promoter"  
 FT terminator 6818..7050  
 FT /\*tag= f  
 FT /\*note= "bovine growth hormone polyA sequence"

XX WO9806437-A2.

XX 19-FEB-1998.

XX 13-AUG-1997; 97WO-US14465.

XX 13-AUG-1996; 96US-0023867.

XX (CHIR ) CHIRON CORP.

XX Cohen F, Dubois-Stringfellow N, Dwarki V, Innis MA;  
 PI Murphy JE, Tetsuo U, Zukermann R;

XX WPI; 1998-159296/14.

XX Polyclonal agents based on alpha-amino acids, able to complex  
 PT with nucleic acid - to facilitate its entry into cell, condense it  
 PT and protect it against serum degradation, particularly for use in  
 PT gene therapy

XX Disclosure; Page 77-80; 100pp; English.

XX This polynucleotide comprises the DNA sequence of vector plasmid

CC PCMVkMTR-EPI, which contains an Epstein-Barr virus (EBV) origin  
 CC of replication from plasmid pCEP4, a coding region for EBV nuclear  
 CC antigen A from pCEP4, a pair of inverted terminal repeats from  
 CC adeno-associated virus, a cytomegalovirus enhancer/promoter, a  
 CC bovine growth hormone polyA sequence, and a kanamycin resistance  
 CC selectable marker. Polynucleotides encoding polypeptides, such as  
 CC erythropoietin or leptin, and ribozymes and antisense  
 CC polynucleotides can be inserted into the vector. The vector is  
 CC preferred for use in novel compositions and methods for improved  
 CC polynucleotide delivery into cells. In these methods, polycationic  
 CC agents are used to increase the frequency of uptake of a  
 CC nucleic acid (see also AA21684-86) into a cell. The polycationic  
 CC agent can condense with the nucleic acid and inhibit serum and/or  
 CC nuclease degradation of the nucleic acid. The nucleic acid can be  
 CC a vector, may express a therapeutic protein or a vaccinating viral  
 CC ribozyme). The methods and compositions can be used in the gene  
 CC therapy of many diseases.  
 XX  
 SQ Sequence 9600 BP; 2326 A; 2376 C; 2817 G; 2081 T; 0 other;

Query Match 100.0%; Score 30; DB 19; Length 9600;  
 Best Local Similarity 100.0%; Pred. No. 0.0023;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgtctcccttgggaatgcccctgagacc 30  
 |||||  
 Db 2041 cgtctcccttgggaatgcccctgagacc 2070

RESULT 14  
 AAA53873  
 ID AAA53873 standard; DNA; 9725 BP.

XX  
 AC AAA53873;  
 XX  
 DT 03-JAN-2001 (first entry)  
 XX  
 DE Expression vector pRIG19.  
 XX  
 KW Vector; endogenous gene; activation; over-expression;  
 KW erythropoietin; growth hormone; drug discovery;  
 KW granulocyte colony stimulating factor; ds.

XX  
 OS Synthetic.  
 XX  
 PN WO200049162-A2.  
 XX  
 PD 24-AUG-2000.

XX  
 PF 22-FEB-2000; 2000WO-US04429.  
 XX  
 PR 19-FEB-1999; 99US-0253022.  
 PR 08-MAR-1999; 99US-0263814.  
 PR 26-MAR-1999; 99US-0276820.

XX  
 PA (ATHE-) ATHERSYS INC.  
 XX  
 PI Harrington JJ, Sherf B, Rundlett S;  
 XX  
 DR WPI; 2000-549276/50.

XX Non-targeted activation of endogenous genes, e.g. for the production of  
 PT erythropoietin, growth hormone or granulocyte-colony stimulating factor  
 PT proteins and for drug discovery

XX Disclosure; Fig 30; 240pp; English.

XX New methods, vectors and cells are described for non-targeted  
 CC activation and over-expression of endogenous genes. The vector  
 CC constructs comprise transcriptional regulatory sequences (TRS) and  
 CC unpaired splice donor sequences (USDS), preferably the vectors

CC comprise (in sequential order) a TRS, an USDS, a rare cutting  
 CC restriction site (RCRS) and a linearization site (LS) with a second  
 CC TRS linked to a selectable marker (SM) lacking a polyadenylation  
 CC signal. The methods, vectors and cells comprising the vectors may  
 CC be used for the non-targeted activation and over-expression of  
 CC endogenous genes, e.g. for the production of proteins (including  
 CC erythropoietin, growth hormone or granulocyte-colony stimulating  
 CC factor) and drug discovery. The advantage of these methods are that  
 CC endogenous genes including those associated with human disease and  
 CC development, may be activated and isolated without prior knowledge  
 CC of the sequence structure, function or expression profile of  
 CC the genes being known.

XX  
 SQ Sequence 9725 BP; 2425 A; 2507 C; 2448 G; 2343 T; 2 other;

Query Match 100.0%; Score 30; DB 21; Length 9725;  
 Best Local Similarity 100.0%; Pred. No. 0.0023;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 cgtctcccttgggaatgcccctgagacc 30  
 |||||  
 Db 3262 cgtctcccttgggaatgcccctgagacc 3291

RESULT 15  
 AAA53879  
 ID AAA53879 standard; DNA; 9732 BP.

XX  
 AC AAA53879;  
 XX  
 DT 03-JAN-2001 (first entry)  
 XX  
 DE Expression vector pRIG-TP.  
 XX  
 KW Vector; endogenous gene; activation; over-expression;  
 KW erythropoietin; growth hormone; drug discovery;  
 KW granulocyte colony stimulating factor; ds.

XX  
 OS Synthetic.  
 XX  
 PN WO200049162-A2.  
 XX  
 PD 24-AUG-2000.

XX  
 PF 22-FEB-2000; 2000WO-US04429.  
 XX  
 PR 19-FEB-1999; 99US-0253022.  
 PR 08-MAR-1999; 99US-0263814.  
 PR 26-MAR-1999; 99US-0276820.

XX  
 PA (ATHE-) ATHERSYS INC.  
 XX  
 PI Harrington JJ, Sherf B, Rundlett S;  
 XX  
 DR WPI; 2000-549276/50.

XX Non-targeted activation of endogenous genes, e.g. for the production of  
 PT erythropoietin, growth hormone or granulocyte colony stimulating factor  
 PT proteins and for drug discovery

XX Example 15; Fig 37; 240pp; English.

XX New methods, vectors and cells are described for non-targeted  
 CC activation and over-expression of endogenous genes. The vector  
 CC constructs comprise transcriptional regulatory sequences (TRS) and  
 CC unpaired splice donor sequences (USDS), preferably the vectors  
 CC comprise (in sequential order) a TRS, an USDS, a rare cutting  
 CC restriction site (RCRS) and a linearization site (LS) with a second  
 CC TRS linked to a selectable marker (SM) lacking a polyadenylation  
 CC signal. The methods, vectors and cells comprising the vectors may  
 CC be used for the non-targeted activation and over-expression of  
 CC endogenous genes, e.g. for the production of proteins (including

CC erythropoietin, growth hormone or granulocyte-colony stimulating  
CC factor) and drug discovery. The advantage of these methods are that  
CC endogenous genes including those associated with human disease and  
CC development, may be activated and isolated without prior knowledge  
CC of the sequence structure, function or expression profile of  
XX the genes being known.

SQ Sequence 9732 BP; 2427 A; 2512 C; 2443 G; 2348 T; 2 other;

Query Match 100.0%; Score 30; DB 21; Length 9732;  
Best Local Similarity 100.0%; Pred. No. 0.0023;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgtctcccccttttggaatggccccctggaccc 30  
|||||  
DB 3270 cgtctcccccttttggaatggccccctggaccc 3299

Search completed: May 18, 2002, 17:45:55  
Job time: 13744 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:46:03 ; Search time 1024.22 Seconds  
(without alignments)  
50.289 Million cell updates/sec

Title: US-09-623-329-26

Perfect score: 30

Sequence: 1 cgggttaactggccaggagagagca 30

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

1:	/SID55/qcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SID55/qcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SID55/qcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SID55/qcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SID55/qcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SID55/qcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SID55/qcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SID55/qcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SID55/qcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SID55/qcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SID55/qcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SID55/qcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SID55/qcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SID55/qcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SID55/qcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SID55/qcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SID55/qcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SID55/qcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SID55/qcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SID55/qcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SID55/qcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SID55/qcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SID55/qcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SID55/qcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	100.0	30	AAZ11676	EBV BARF-1 specifi
2	20	66.7	1389	AAV29992	Human BY55 cDNA se
3	19.6	65.3	2007	AAK73932	Human immune/haema
4	19.6	65.3	2007	AAK73933	Human immune/haema
5	19.6	65.3	22680	AAK66308	Human immune/haema
6	19.6	65.3	22680	AAK73334	Human immune/haema
7	19.6	65.3	22680	AAK73344	Human immune/haema
8	19.6	65.3	22680	AAK73625	Human immune/haema
9	19.6	65.3	22680	AAK73847	Human immune/haema

10	19.6	65.3	22680	22	AAK73934	Human immune/haema
11	19.6	65.3	22680	22	AAK78350	Human immune/haema
c 12	19.4	64.7	27804	22	AAK86476	Human immune/haema
c 13	19.4	64.7	34794	22	AAK88884	Ad35 nucleotide se
c 14	18.8	62.7	623	22	AAK92295	Human cDNA 5'-end
c 15	18.8	62.7	623	22	AAK93820	Human cDNA clone r
c 16	18.8	62.7	863	22	AAF27732	Human transport pr
c 17	18.8	62.7	1583	22	AAH15123	Human cDNA sequenc
c 18	18.8	62.7	1642	21	AAF21777	Human breast and o
c 19	18.8	62.7	2038	22	AAO5523	Human reproductive
c 20	18.8	62.7	2038	22	AAK81824	Human immune/haema
c 21	18.8	62.7	2047	22	AAK62145	Human immune/haema
c 22	18.8	62.7	2184	22	AAK94579	Human full-length
c 23	18.8	62.7	2331	23	AAK87248	DNA encoding novel
c 24	18.8	62.7	3471	22	AAH15559	Human cDNA sequenc
c 25	18.6	62.0	2741	22	ABA20115	Human nervous syst
c 26	18.6	62.0	2997	20	AA228284	Rat neuronal immed
c 27	18.6	62.0	2997	22	AAI66277	Rat apoptase L100
c 28	18.4	61.3	561	22	AAK57923	Human immune/haema
c 29	18.4	61.3	945	23	AAK83268	DNA encoding novel
c 30	18.4	61.3	7405	22	AAK83850	Human immune/haema
c 31	18.4	61.3	8165	22	AAK82672	Human immune/haema
c 32	18.2	60.7	1028	23	AAK81438	DNA encoding novel
c 33	18.2	60.7	3196	22	AAV20420	Human discs large
c 34	18	60.0	567	22	AAJ16657	Human novel protei
c 35	18	60.0	567	22	AAK87993	Human digestive sy
c 36	18	60.0	1047	22	ABA11221	Human nervous syst
c 37	18	60.0	2054	21	AAZ43779	Human fetal brain
c 38	18	60.0	4298	22	ABA14784	Human nervous syst
c 39	18	60.0	29163	22	AAO5121	Human reproductive
c 40	17.8	59.3	877	22	ABA18230	Human nervous syst
c 41	17.8	59.3	1732	21	AAZ98218	Human signal pepti
c 42	17.8	59.3	2085	21	AAK77828	Human cancer assoc
c 43	17.8	59.3	2249	22	AAK72758	Human secreted pro
c 44	17.8	59.3	2379	20	AAZ20859	Polynucleotide seq
c 45	17.8	59.3	2379	22	AAZ59296	Human cDNA encodin

#### ALIGNMENTS

RESULT 1

AAZ11676

ID AAZ11676 standard; DNA; 30 BP.

XX

AC AAZ11676;

XX

DT 19-NOV-1999 (first entry)

XX

DE EBV BARF-1 specific probe.

XX

Epstein Barr Virus; EBV infection; viral; gene transcription; EBER-1;

KW Epstein Barr early RNA; Epstein Barr nuclear antigen 1; EBNA-1; LMP-1;

KW latent membrane protein; LMP-2; vIL10; BCRF-1; BARF1; BDLF2; NASBA;

KW EBV-associated malignancy; probe; ss.

XX

OS Synthetic.

OS Epstein-barr virus.

XX

PN WO9945155-A2.

XX

PD 10-SEP-1999.

XX

PF 01-MAR-1999; 99WO-EP01392.

XX

PR 04-MAR-1998; 98EP-0200655.

PR 14-DEC-1998; 98EP-0204231.

XX

PA (ALKU ) AKZO NOBEL NV.

XX

PI Vervoort MBHJ, Van Den Brule AJC, Middeldorp JM;

XX

WPI; 1999-551051/46.

```
XX Identifying Epstein Barr Virus infection -
XX Claim 14; Page 22; 50pp; English.
XX
CC The invention provides methods for identifying an Epstein Barr Virus
CC (EBV) infection, that comprises determining viral gene transcription
CC patterns by amplification of specific RNA sequences. The binding sites
CC of the oligos suitable for amplification are located in the following
CC genes: Epstein Barr early RNA (EBER-1), Epstein Barr nuclear antigen 1
CC (EBNA-1), latent membrane protein 1 (LMP-1), LMP-2, and vIL10 (BCRF-1),
CC BARF1 and BDLF2. The method comprises (a) amplifying a target sequence
CC within one or more RNA(s) transcribed from above gene sequences and the
CC (b) detecting the amplified products, determining the transcription
CC pattern and identifying the corresponding EBV-associated malignancy. The
CC RNA is amplified using a transcription based amplification technique such
CC as NASBA. The invention is used to diagnose malignant and non-malignant
CC EBV-associated diseases. The present sequence represents a probe specific
CC for BARF-1 RNA comprising a detectable label.
XX
SQ Sequence 30 BP; 8 A; 6 C; 11 G; 5 T; 0 other;

Query Match 100.0%; Score 30; DB 20; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctggtttaaactggccaggagagagca 30
   |||||
DB 1 ctggtttaaactggccaggagagagca 30

RESULT 2
ID AAV29992 standard; cDNA; 1389 BP.
XX
AC AAV29992;
XX
DT 14-SEP-1998 (first entry)
XX
DE Human BY55 cDNA sequence.
XX
KW BY55; COS cell; polyacrylamide gel electrophoresis; cysteine residue;
KW disulfide bond; multimeric; immune modulation; cytolytic activity; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 216..761
FT /*tag= a
FT /product= "BY55 protein"
FT exon 1..37
FT /*tag= b
FT /number= 1
FT /note= "may be a partial exon; 5' end of the exon is
FT not defined in the specification"
FT intron 38..143
FT /*tag= b
FT /number= 1
FT /cons_splice= (5'site: NO, 3'site: YES)
FT /note= "this intronic sequence is found in one of
FT the BY55 clones isolated"
FT exon 144..1389
FT /*tag= c
FT /number= 2
FT /note= "may be a partial exon; 3' end of exon is not
FT defined in the specification"
FT sig_peptide 216..290
FT /*tag= d
FT mat_peptide 291..758
FT /*tag= e
FT polyA_signal 1341..1347
FT /*tag= f
```

```
XX WO9821240-A1.
XX
XX 22-MAY-1998.
XX
XX 12-NOV-1997; 97WO-US20602.
XX
XX 13-NOV-1996; 96US-0747853.
XX
XX 12-NOV-1996; 96US-0030527.
XX
XX (DAND ) DANA FARBER CANCER INST INC.
XX
XX Anumanthan A, Freeman GJ;
XX WPI; 1998-297867/26.
XX
XX P-PSDB; AAW56618.
XX
XX Novel immune associated protein BY55 - useful for screening for
XX compounds to modulate immune responses
XX
XX Claim 2; Fig 2A; 69pp; English.
XX
XX This nucleic acid sequence is the human BY55 cDNA, which encodes a human
XX BY55 protein that can be expressed in eukaryotic cells, (such as COS
XX cells). The molecular weight of this recombinant protein was determined
XX by polyacrylamide gel electrophoresis as 80 kD. It is also seen that the
XX cDNA encodes a peptide that contains six cysteine residues, and would be
XX expected to have a nonprocessed molecular weight of 24 kD. Due to the
XX interchain disulfide bonds formed by these six cysteine residues in the
XX mature protein, it appears that the protein contains multiple chains,
XX thus is in a multimeric form. It was also found that a single
XX recombinant chain could be obtained. The protein produced from the BY55
XX cDNA can be used to both screen for compounds that modulate or inhibit
XX the activity of the BY55 expressing cells, or it can be used to modulate
XX the activity of BY55 expressing cells. The compounds are useful for
XX controlling the immune response and cytolytic activity of BY55 expressing
XX cells.
XX
XX Sequence 1389 BP; 451 A; 310 C; 309 G; 319 T; 0 other;

Query Match 66.7%; Score 20; DB 19; Length 1389;
Best Local Similarity 82.1%; Pred. No. 49;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ctggtttaaactggccaggagagagca 28
   ||| ||||| ||||| ||||| ||
DB 443 CTGTTTAAACTGGTCTCAGGAGACAG 416

RESULT 3
ID AAK73932 standard; DNA; 2007 BP.
XX
XX AAK73932;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:28744.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX 04-FEB-2000; 2000US-0180628.
```

PR	24-FEB-2000;	2000US-0184664.
PR	02-MAR-2000;	2000US-0186350.
PR	16-MAR-2000;	2000US-0198374.
PR	17-MAR-2000;	2000US-0190077.
PR	18-APR-2000;	2000US-0198123.
PR	19-MAY-2000;	2000US-0205515.
PR	07-JUN-2000;	2000US-0209467.
PR	28-JUN-2000;	2000US-0214886.
PR	30-JUN-2000;	2000US-0215135.
PR	07-JUL-2000;	2000US-0216647.
PR	07-JUL-2000;	2000US-0216880.
PR	11-JUL-2000;	2000US-0217487.
PR	11-JUL-2000;	2000US-0217496.
PR	14-JUL-2000;	2000US-0218290.
PR	26-JUL-2000;	2000US-0220963.
PR	26-JUL-2000;	2000US-0220984.
PR	14-AUG-2000;	2000US-0224518.
PR	14-AUG-2000;	2000US-0224519.
PR	14-AUG-2000;	2000US-0225213.
PR	14-AUG-2000;	2000US-0225214.
PR	14-AUG-2000;	2000US-0225266.
PR	14-AUG-2000;	2000US-0225267.
PR	14-AUG-2000;	2000US-0225268.
PR	14-AUG-2000;	2000US-0225270.
PR	14-AUG-2000;	2000US-0225447.
PR	14-AUG-2000;	2000US-0225577.
PR	14-AUG-2000;	2000US-0225578.
PR	14-AUG-2000;	2000US-0225759.
PR	18-AUG-2000;	2000US-0226759.
PR	22-AUG-2000;	2000US-0226681.
PR	22-AUG-2000;	2000US-0226688.
PR	22-AUG-2000;	2000US-0226691.
PR	23-AUG-2000;	2000US-0227009.
PR	30-AUG-2000;	2000US-0228924.
PR	01-SEP-2000;	2000US-0229287.
PR	01-SEP-2000;	2000US-0229343.
PR	01-SEP-2000;	2000US-0229344.
PR	01-SEP-2000;	2000US-0229345.
PR	05-SEP-2000;	2000US-0229509.
PR	05-SEP-2000;	2000US-0229513.
PR	06-SEP-2000;	2000US-0230437.
PR	06-SEP-2000;	2000US-0230438.
PR	08-SEP-2000;	2000US-0231242.
PR	08-SEP-2000;	2000US-0231243.
PR	08-SEP-2000;	2000US-0231244.
PR	08-SEP-2000;	2000US-0231413.
PR	08-SEP-2000;	2000US-0231414.
PR	08-SEP-2000;	2000US-0232080.
PR	08-SEP-2000;	2000US-0232081.
PR	12-SEP-2000;	2000US-0231968.
PR	14-SEP-2000;	2000US-0232397.
PR	14-SEP-2000;	2000US-0232398.
PR	14-SEP-2000;	2000US-0232399.
PR	14-SEP-2000;	2000US-0232400.
PR	14-SEP-2000;	2000US-0232401.
PR	14-SEP-2000;	2000US-0233063.
PR	14-SEP-2000;	2000US-0233064.
PR	14-SEP-2000;	2000US-0233065.
PR	21-SEP-2000;	2000US-0234223.
PR	21-SEP-2000;	2000US-0234274.
PR	25-SEP-2000;	2000US-0234499.
PR	25-SEP-2000;	2000US-0234998.
PR	26-SEP-2000;	2000US-0235484.
PR	27-SEP-2000;	2000US-0235834.
PR	27-SEP-2000;	2000US-0235836.
PR	29-SEP-2000;	2000US-0236327.
PR	29-SEP-2000;	2000US-0236367.
PR	29-SEP-2000;	2000US-0236368.
PR	29-SEP-2000;	2000US-0236369.
PR	02-OCT-2000;	2000US-0236670.
PR	02-OCT-2000;	2000US-0236682.
PR	02-OCT-2000;	2000US-0237037.
PR	02-OCT-2000;	2000US-0237038.



SQ Sequence 2007 BP; 402 A; 608 C; 570 G; 427 T; 0 other;

Query Match 65.3%; Score 19.6; DB 22; Length 2007;  
Best Local Similarity 84.6%; Pred. No. 75;  
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps

OY 3 gqlltaactlggcccgagaggag 28  
| | | | | | | | | | | | | |  
Db 86 gctcggacctggcccgagaggag 111

RESULT 5  
AAK66308  
ID AAK66308 Standard; DNA: 22680 bp.  
XX  
AC AAK66308;  
XX  
DT 06-NOV-2001 (first entry)  
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:21120.  
XX Human; immune: haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX Homo sapiens.  
XX WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01354.  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.

(HUMA-) HUMAN GENOME SCI INC.  
  
Rosen CA, Batash SC, Ruben SM;  
WPI; 2001-483426/52.  
  
Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
useful for preventing, diagnosing and/or treating cancers and  
metastasis -  
  
Disclosure; SEQ ID NO 28745; 3071pp + Sequence listing; English.  
  
AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
activity, and can be used in gene therapy and vaccine production. (I)  
proteins and polynucleotides may be used in the prevention, diagnosis and  
treatment of diseases associated with inappropriate (I) expression. For  
example, they may be used to treat disorders associated with decreased  
expression by rectifying mutations or deletions in a patient's genome  
that affect the activity of (I) by expressing inactive proteins or to  
supplement the patients own production of (I). Additionally, (I)  
polynucleotides may be used to produce the secreted (I), by inserting the  
nucleic acids into a host cell and culturing the cell to express the  
protein. (I) proteins and polynucleotides may be used to prevent,  
diagnose and treat immune/haematopoietic-related diseases, especially  
cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
to AAK87694 represent human immune/haematopoietic antigen genomic  
sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
represent sequences used in the exemplification of the present invention.



PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 06-SEP-2000; 2000US-0230439.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 13-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.

PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-483426/52.  
 DR Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis.  
 XX Disclosure; SEQ ID NO 21120; 3071pp + Sequence Listing; English.  
 PS AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting the  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/haematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/haematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
 CC represent sequences used in the exemplification of the present invention.  
 XX Sequence 22680 BP; 4444 A; 6635 C; 7046 G; 4555 T; 0 other;  
 SQ  
 Query Match 65.3%; Score 19.6; DB 22; Length 22680;  
 Best Local Similarity 84.6%; Pred. No. 95;  
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 3 ggttaactggggccaggagagag 28  
 Db 7948 gctctggactggggccaggagag 7973  
 RESULT 6  
 ID AAK73334 standard; DNA; 22680 BP.  
 XX AAK73334;  
 XX 06-NOV-2001 (first entry)

XX	Human immune/haematopoietic antigen genomic sequence	SEQ ID NO:28146.
DE	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;	
XX	cytostatic; gene therapy; vaccine; metastasis; ds.	
KW	Homo sapiens.	
KW	WO200157182-A2.	
XX	09-AUG-2001.	
XX	17-JAN-2001; 2001WO-USO1354.	
PF	31-JAN-2000; 2000US-0179065.	
XX	04-FEB-2000; 2000US-0180628.	
XX	24-FEB-2000; 2000US-0184664.	
XX	02-MAR-2000; 2000US-0186350.	
XX	16-MAR-2000; 2000US-0189874.	
XX	17-MAR-2000; 2000US-0190076.	
XX	18-APR-2000; 2000US-0198123.	
XX	19-MAY-2000; 2000US-0205515.	
XX	07-JUN-2000; 2000US-0209467.	
XX	28-JUN-2000; 2000US-0214886.	
XX	30-JUN-2000; 2000US-0215135.	
XX	07-JUL-2000; 2000US-0216647.	
XX	07-JUL-2000; 2000US-0216880.	
XX	11-JUL-2000; 2000US-0217487.	
XX	11-JUL-2000; 2000US-0217496.	
XX	14-JUL-2000; 2000US-0218290.	
XX	26-JUL-2000; 2000US-0220963.	
XX	26-JUL-2000; 2000US-0220964.	
XX	14-AUG-2000; 2000US-0224518.	
XX	14-AUG-2000; 2000US-0224519.	
XX	14-AUG-2000; 2000US-0225213.	
XX	14-AUG-2000; 2000US-0225214.	
XX	14-AUG-2000; 2000US-0225266.	
XX	14-AUG-2000; 2000US-0225267.	
XX	14-AUG-2000; 2000US-0225268.	
XX	14-AUG-2000; 2000US-0225270.	
XX	14-AUG-2000; 2000US-0225447.	
XX	14-AUG-2000; 2000US-0225757.	
XX	14-AUG-2000; 2000US-0225758.	
XX	14-AUG-2000; 2000US-0225759.	
XX	18-AUG-2000; 2000US-0226279.	
XX	22-AUG-2000; 2000US-0226681.	
XX	22-AUG-2000; 2000US-0226868.	
XX	22-AUG-2000; 2000US-0227182.	
XX	23-AUG-2000; 2000US-0227009.	
XX	30-AUG-2000; 2000US-0228924.	
XX	01-SEP-2000; 2000US-0229287.	
XX	01-SEP-2000; 2000US-0229343.	
XX	01-SEP-2000; 2000US-0229344.	
XX	01-SEP-2000; 2000US-0229345.	
XX	05-SEP-2000; 2000US-0229509.	
XX	05-SEP-2000; 2000US-0229513.	
XX	06-SEP-2000; 2000US-0230437.	
XX	06-SEP-2000; 2000US-0230438.	
XX	08-SEP-2000; 2000US-0231242.	
XX	08-SEP-2000; 2000US-0231243.	
XX	08-SEP-2000; 2000US-0231244.	
XX	08-SEP-2000; 2000US-0231413.	
XX	08-SEP-2000; 2000US-0231414.	
XX	08-SEP-2000; 2000US-0232080.	
XX	08-SEP-2000; 2000US-0232081.	
XX	12-SEP-2000; 2000US-0231968.	
XX	14-SEP-2000; 2000US-0232397.	
XX	14-SEP-2000; 2000US-0232398.	
XX	14-SEP-2000; 2000US-0232399.	
XX	14-SEP-2000; 2000US-0232400.	
XX	14-SEP-2000; 2000US-0232401.	
XX	14-SEP-2000; 2000US-0233063.	
XX	14-SEP-2000; 2000US-0233064.	

	PR	08-DEC-2000; 2000US-0251989.		PR	17-MAR-2000; 2000US-0190076
	PR	08-DEC-2000; 2000US-0251990.		PR	18-APR-2000; 2000US-0198123
	PR	11-DEC-2000; 2000US-0254097.		PR	19-MAY-2000; 2000US-0205515
	PR	05-JAN-2001; 2001US-0259678.		PR	07-JUN-2000; 2000US-0209467
XX	XX	(HUMA-) HUMAN GENOME SCI INC.		PR	28-JUN-2000; 2000US-0214886
XX	XX	Rosen CA, Barash SC, Ruben SM;		PR	30-JUN-2000; 2000US-0215135
XX	PI	WPI; 2001-483426/52.		PR	07-JUL-2000; 2000US-0216647
XX	DR			PR	11-JUL-2000; 2000US-0216880
PP	PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,		PR	11-JUL-2000; 2000US-0217487
PP	PT	useful for preventing, diagnosing and/or treating cancers and		PR	14-JUL-2000; 2000US-0217496
PT	PT	metastasis .		PR	26-JUL-2000; 2000US-0220963
XX	XX	Disclosure; SEQ ID NO 28146; 3071pp + Sequence Listing; English.		PR	26-JUL-2000; 2000US-0220964
PS	PS	AAK54951 to AAK64702 encode the human Immune/haematopoietic antigen (I)		PR	14-AUG-2000; 2000US-0224518
XX	CC	amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic		PR	14-AUG-2000; 2000US-0224519
CC	CC	activity, and can be used in gene therapy and vaccine production. (I)		PR	14-AUG-2000; 2000US-0225213
CC	CC	proteins, and polynucleotides may be used in the prevention, diagnosis and		PR	14-AUG-2000; 2000US-0225214
CC	CC	treatment of diseases associated with inappropriate (I) expression. For		PR	14-AUG-2000; 2000US-0225267
CC	CC	example, they may be used to treat disorders associated with decreased		PR	14-AUG-2000; 2000US-0225268
CC	CC	expression by rectifying mutations or deletions in a patient's genome		PR	14-AUG-2000; 2000US-0225270
CC	CC	that affect the activity of (I) by expressing inactive proteins or to		PR	14-AUG-2000; 2000US-0225447
CC	CC	supplement the patients own production of (I). Additionally, (I)		PR	14-AUG-2000; 2000US-0225557
CC	CC	polynucleotides may be used to produce the secreted (I), by inserting		PR	14-AUG-2000; 2000US-0225758
CC	CC	the nucleic acids into a host cell and culturing the cell to express the		PR	14-AUG-2000; 2000US-0225759
CC	CC	protein. (I) proteins and polynucleotides may be used to prevent,		PR	18-AUG-2000; 2000US-0226279
CC	CC	diagnose and treat immune/haematopoietic-related diseases, especially		PR	22-AUG-2000; 2000US-0226681
CC	CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703		PR	22-AUG-2000; 2000US-0227182
CC	CC	to AAK87694 represent human immune/haematopoietic antigen genomic		PR	23-AUG-2000; 2000US-0227009
CC	CC	sequences from the present invention. AAK54942 to AAK54950 and AAM82169		PR	30-AUG-2000; 2000US-0228924
CC	CC	represent sequences used in the exemplification of the present invention.		PR	01-SEP-2000; 2000US-0229287
XX	XX	Sequence 22680 BP; 4444 A; 6635 C; 7046 G; 4555 T; 0 other:		PR	01-SEP-2000; 2000US-0229343
SQ				PR	01-SEP-2000; 2000US-0229344
				PR	01-SEP-2000; 2000US-0229345
				PR	05-SEP-2000; 2000US-0229509
				PR	05-SEP-2000; 2000US-0229513
				PR	06-SEP-2000; 2000US-0230437
				PR	06-SEP-2000; 2000US-0230438
				PR	08-SEP-2000; 2000US-0231242
				PR	08-SEP-2000; 2000US-0231243
				PR	08-SEP-2000; 2000US-0231244
				PR	08-SEP-2000; 2000US-0231413
				PR	08-SEP-2000; 2000US-0231414
				PR	08-SEP-2000; 2000US-0232080
				PR	08-SEP-2000; 2000US-0232081
				PR	12-SEP-2000; 2000US-0231968
				PR	14-SEP-2000; 2000US-0232397
				PR	14-SEP-2000; 2000US-0232398
				PR	14-SEP-2000; 2000US-0232399
				PR	14-SEP-2000; 2000US-0232400
				PR	14-SEP-2000; 2000US-0232401
				PR	14-SEP-2000; 2000US-0233063
				PR	14-SEP-2000; 2000US-0233064
				PR	14-SEP-2000; 2000US-0233065
				PR	21-SEP-2000; 2000US-0234223
				PR	21-SEP-2000; 2000US-0234274
				PR	25-SEP-2000; 2000US-0234997
				PR	25-SEP-2000; 2000US-0234998
				PR	26-SEP-2000; 2000US-0235484
				PR	27-SEP-2000; 2000US-0235834
				PR	27-SEP-2000; 2000US-0235836
				PR	29-SEP-2000; 2000US-0236327
				PR	29-SEP-2000; 2000US-0236367



PR	14-AUG-2000;	2000US-02252568	PR
PR	14-AUG-2000;	2000US-02252570	PR
PR	14-AUG-2000;	2000US-02252547	PR
PR	14-AUG-2000;	2000US-02252577	PR
PR	14-AUG-2000;	2000US-02252558	PR
PR	14-AUG-2000;	2000US-02252759	PR
PR	18-AUG-2000;	2000US-02262799	PR
PR	18-AUG-2000;	2000US-02262759	PR
PR	22-AUG-2000;	2000US-02266681	PR
PR	22-AUG-2000;	2000US-02266686	PR
PR	22-AUG-2000;	2000US-0227182	PR
PR	23-AUG-2000;	2000US-02270009	PR
PR	30-AUG-2000;	2000US-0228924	PR
PR	01-SEP-2000;	2000US-0229287	PR
PR	01-SEP-2000;	2000US-0229343	PR
PR	01-SEP-2000;	2000US-0229344	PR
PR	01-SEP-2000;	2000US-0229345	PR
PR	05-SEP-2000;	2000US-0229509	PR
PR	05-SEP-2000;	2000US-0229513	PR
PR	06-SEP-2000;	2000US-0230437	PR
PR	08-SEP-2000;	2000US-0230438	PR
PR	08-SEP-2000;	2000US-0231242	PR
PR	08-SEP-2000;	2000US-0231243	PR
PR	08-SEP-2000;	2000US-0231244	PR
PR	08-SEP-2000;	2000US-0231413	PR
PR	08-SEP-2000;	2000US-0231414	PR
PR	08-SEP-2000;	2000US-0232080	PR
PR	08-SEP-2000;	2000US-0232081	PR
PR	12-SEP-2000;	2000US-0231968	PR
PR	12-SEP-2000;	2000US-0232397	PR
PR	14-SEP-2000;	2000US-0232398	PR
PR	14-SEP-2000;	2000US-0232399	PR
PR	14-SEP-2000;	2000US-0232400	PR
PR	14-SEP-2000;	2000US-0232401	PR
PR	14-SEP-2000;	2000US-0232402	PR
PR	14-SEP-2000;	2000US-0233063	PR
PR	14-SEP-2000;	2000US-0233064	PR
PR	14-SEP-2000;	2000US-0233065	PR
PR	21-SEP-2000;	2000US-0234223	PR
PR	21-SEP-2000;	2000US-0234274	PR
PR	25-SEP-2000;	2000US-0234997	PR
PR	25-SEP-2000;	2000US-0234998	PR
PR	26-SEP-2000;	2000US-0235484	PR
PR	27-SEP-2000;	2000US-0235834	PR
PR	27-SEP-2000;	2000US-0235836	PR
PR	29-SEP-2000;	2000US-0236327	PR
PR	29-SEP-2000;	2000US-0236367	PR
PR	29-SEP-2000;	2000US-0236368	PR
PR	29-SEP-2000;	2000US-0236369	PR
PR	29-SEP-2000;	2000US-0236370	PR
PR	02-OCT-2000;	2000US-0236802	PR
PR	02-OCT-2000;	2000US-0237037	PR
PR	02-OCT-2000;	2000US-0237038	PR
PR	02-OCT-2000;	2000US-0237039	PR
PR	13-OCT-2000;	2000US-0237040	PR
PR	13-OCT-2000;	2000US-0239935	PR
PR	13-OCT-2000;	2000US-0239937	PR
PR	20-OCT-2000;	2000US-0239937	PR
PR	20-OCT-2000;	2000US-0240960	PR
PR	20-OCT-2000;	2000US-0241221	PR
PR	20-OCT-2000;	2000US-0241785	PR
PR	20-OCT-2000;	2000US-0241786	PR
PR	20-OCT-2000;	2000US-0241787	PR
PR	20-OCT-2000;	2000US-0241808	PR
PR	20-OCT-2000;	2000US-0241809	PR
PR	20-OCT-2000;	2000US-0241826	PR
PR	01-NOV-2000;	2000US-0244617	PR
PR	08-NOV-2000;	2000US-0246474	PR
PR	08-NOV-2000;	2000US-0246475	PR
PR	08-NOV-2000;	2000US-0246476	PR
PR	08-NOV-2000;	2000US-0246477	PR
PR	08-NOV-2000;	2000US-0246478	PR
PR	08-NOV-2000;	2000US-0246523	PR
PR	08-NOV-2000;	2000US-0246524	PR
PR	08-NOV-2000;	2000US-0246525	PR
PR	08-NOV-2000;	2000US-0246526	PR





25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 13-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 20-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249266.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX PA  
 XX PI  
 XX Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-483426/52.  
 XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis.  
 XX Disclosure; SEQ ID NO 28746; 3071pp + Sequence Listing; English.  
 PS  
 XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/hematopoietic-related diseases, especially  
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/hematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
 CC represent sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 22680 BP; 4444 A; 6635 C; 7046 G; 4555 T; 0 other;  
 Query Match 65.3%; Score 19.6; DB 22; Length 22680;  
 Best Local Similarity 84.6%; Pred. No. 95;  
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 3 ggtttaaactggcccgaggagagag 28  
 Db 7948 gctctgactggcccgaggagagag 7973  
 RESULT 11  
 AAK78350  
 ID AAK78350 standard; DNA; 22680 BP.  
 AC AAK78350;  
 XX  
 XX 07-NOV-2001 (first entry)  
 XX Human immune/hematopoietic antigen genomic sequence SEQ ID NO:33162.  
 XX Human; immune; haematopoietic; immune/hematopoietic antigen; cancer;  
 KW cytostatic; gene therapy; vaccine; metastasis; ds.  
 XX Homo sapiens.  
 XX WO200157182-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 XX 17-JAN-2001; 2001WO-US01354.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.



PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226686.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 11-DEC-2000; 2000US-0251990.  
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Kuben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -

Disclosure; SEQ ID NO 33162; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 14:58:27 ; Search time 2878.96 Seconds  
(without alignments)  
145.376 Million cell updates/sec

Title: US-09-623-329-23

Perfect score: 20  
Sequence: 1 ggctgtcaccgtttttcttgg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pi.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htgo\_inv.\*

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----					

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

1	20	100.0	20	6	AX018464
2	20	100.0	172281	14	EBV
3	20	100.0	184113	14	HS48958RAJ
4	18.4	92.0	26104	1	AF005744
5	18.4	92.0	26201	1	AF369954
6	17.4	87.0	91316	2	AC094771
7	17.4	87.0	110000	2	LMFLCHR34_05
8	17.4	87.0	110000	2	LMFLCHR34_06
9	16.8	84.0	566	11	G58484
10	16.8	84.0	622	9	HS4342581
11	16.8	84.0	2876	8	SCYDL096C
12	16.8	84.0	3375	6	A37576
13	16.8	84.0	3375	6	I83618
14	16.8	84.0	3375	8	YSCDPT1A
15	16.8	84.0	13251	3	AF148954
16	16.8	84.0	13761	3	AF148953
17	16.8	84.0	18251	1	AE001791
18	16.8	84.0	28740	9	AL158136
19	16.8	84.0	38516	8	SCCHR1V1A
20	16.8	84.0	39522	9	HSC1170
21	16.8	84.0	41374	2	AC087830
22	16.8	84.0	44699	3	AF040648
23	16.8	84.0	54126	3	AC006834
24	16.8	84.0	74286	2	AC069581
25	16.8	84.0	103616	9	AL138689
26	16.8	84.0	113983	9	HSDJ149L1
27	16.8	84.0	125322	2	AC108017
28	16.8	84.0	134105	5	AC096845
29	16.8	84.0	139399	8	AP002865
30	16.8	84.0	149307	2	AC010858
31	16.8	84.0	150379	8	AP003074
32	16.8	84.0	151623	2	AC009857
33	16.8	84.0	151760	2	AC087593
34	16.8	84.0	152651	2	AC015556
35	16.8	84.0	157439	9	AC026355
36	16.8	84.0	160674	30	AC060811
37	16.8	84.0	162965	2	AC021935
38	16.8	84.0	163292	2	AC104850
39	16.8	84.0	163495	9	AC024183
40	16.8	84.0	163782	9	AC105450
41	16.8	84.0	164364	2	AC099332
42	16.8	84.0	164434	2	AC036216
43	16.8	84.0	164891	2	AC022848
44	16.8	84.0	167429	2	AC010952
45	16.8	84.0	171014	2	AC108489

## ALIGNMENTS

RESULT 1

AX018464

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BASE COUNT

1 a

6 c

6 g

7 t

AX018464 Sequence 20 bp DNA linear PAT 07-SEP-2000  
Sequence 23 from Patent WO945155.  
AX018464  
AX018464.1 GI:10042615

Epstein-Barr virus.  
Human herpesvirus 4

Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Gammaherpesvirinae; Lymphocryptovirus.

1 (bases 1 to 20)

Mideldorp, J.M., Van Den Brule, A.J. and Vervoot, M.B.

Oligonucleotides for the amplification and detection of Epstein  
Barr virus (ebv) nucleic acid

Patent: WO 945155-A 23 10-SEP-1999.

MIDDELDORP JAAP MICHIEL (NL); AKZO NOBEL NV (NL); DEN BRULE

ADRIANUS JOHANNES CH (NL); VERVOORT MARCEL BARTOLINA HEND (NL)

Location/Qualifiers

1..20

/organism="Human herpesvirus 4"

/db\_xref="taxon:10376"

1 a

6 c

6 g

7 t

[illegible]

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

18 (bases 1 to 172281)  
Farrell,P.J. and Barrell,B.G.  
Direct Submission  
Submitted (05-JUN-1984)  
19 (bases 1 to 172281)  
Farrell,P.J.  
Direct Submission  
Submitted (18-MAR-1988) Farrell P., Ludwig Institute for Cancer  
Research, St. Mary's Hospital Medical School, Norfolk Place London  
W2 1PG  
CDS

COMMENT

Listed under this feature are all known protein coding regions as well as all the major open reading frames in the sequence. In general the term major is taken as the longest frame in a particular region taking into account the adjacent longest frames and likely transcription signals. Note that on this basis some long overlapping frames have been excluded and on the other hand some small frames have been included which might represent exons or genes because they occur in a logical combination with other features or because of some other experimental data. The reading frames are named according to the Bam HI fragment in which they start. eg BAUF3 is the third leftward frame starting in Bam HI fragment A. BORF1 is the first rightward frame in Bam HI fragment O. If there is an obvious TATA sequence followed by an in frame Met codon that satisfies the rules of Kozak [12] in that there is a purine at -3 and/or a G at +4 then the reading frame is numbered from the A of the ATG to the base preceding the termination codon. If there is no obvious initiation codon or there is a substantial reading frame in phase before the ATG then the reading frame is numbered from the first base of the first codon.

SITES of POLYA signals  
This feature lists all occurrences of the sequence AATAAA which is found normally approximately 20 bases upstream of the mRNA processing/polyA addition site. The rarely used homolog ATTAAA is only listed when it is found in a position close to the end of a major reading frame.

SITES of DONOR and ACCEPT sequences  
This is not a comprehensive listing of all such sequences and only the positions of a few have been noted because they occur in potentially interesting positions. The number quoted in the table is the position of the terminal base in the intron in each case. Restriction enzyme SITES.  
Only the positions of the sites Bam HI (BAM) are listed.

RPT  
This feature is used to define repetitive sequences.

SITE DEL  
This feature defines deletions in B95-8 with respect to other strains such as RAJ1 and also to deletions in other strains such as P3HR1 and DAUDI with respect to B95-8.

SITE HPN  
Denotes sequences with twofold symmetry ie could form hairpin loops. This is not a comprehensive list - only a few occurrences noted.

ORGRPL  
Denotes the region that encompasses an origin of replication (ori P). [13].  
NUMBERING

The DNA sequence of B95-8 EBV has been revised [19]. The original (Baer et al, 1984) base 359 has been deleted so the new sequence around that position reads TCAGTCTTT. To avoid renumbering the entire sequence, position 1 has been moved 1 base to the left of the EcoRI site separating EcoRI Dhet from EcoRI I (ie the first A of AGAATTC).

FEATURES

source  
1. 172281  
/organism="Human herpesvirus 4"  
/strain="B95-8"  
/db\_xref="taxon:10376"

Query Match 100.0%; Score 20; DB 14; Length 172281;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggctgtcaccgcgtcttcttgg 20  
|||||  
Db 165560 GGTCTCACCGCTTCTTGG 165579

RESULT 3  
HS4B958RAJ

LOCUS  
DEFINITION  
HS4B958RAJ 184113 bp DNA linear VRL 12-APR-1996  
Epstein-Barr virus, artifactual joining of B95-8 complete genome  
and the sequences from Raji of the large deletion found in B95-8.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
M80517 M75989  
M80517.1 GI:330330

ORGANISM  
Human herpesvirus 4 DNA.  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Gammaherpesvirinae; Lymphocryptovirus.

REFERENCE:

AUTHORS  
1 (sites)  
Baer,R.J., Bankier,A.T., Biggin,M.D., Deininger,P.L., Farrell,P.J.,  
Gibson,T.J., Hatfull,G.F., Hudson,G.S., Satchwell,S.C., Seguin,C.,  
Tiffinell,P.S. and Barrell,B.G.

TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
DNA sequence and expression of the B95-8 Epstein-Barr virus genome  
Nature 310 (5974), 207-211 (1984)  
84270667

AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
2 (sites)  
Parker,B.D. Bankier,A., Satchwell,S., Barrell,B. and Farrell,P.J.  
Sequence and transcription of Raji Epstein-Barr virus DNA spanning  
the B95-8 deletion region  
Virology 179 (1), 339-346 (1990)  
91021036

AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
3 (sites)  
Sample,J., Brooks,L., Sample,C., Young,L., Rowe,M., Gregory,C.,  
Rickinson,A. and Kieff,E.  
Restricted Epstein-Barr virus protein expression in Burkitt  
lymphoma is due to a different Epstein-Barr nuclear antigen 1  
transcriptional initiation site  
Proc. Natl. Acad. Sci. U.S.A. 88 (14), 6343-6347 (1991)  
91296817

AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
4 (bases 1 to 184113)  
Jenson,H.B.  
GenBank Curator Program  
Unpublished (1992)

COMMENT  
The B95-8 genome (V01555) has a large deletion in the right side of  
the genome which has been sequenced in Raji (M35547). These  
sequences have been joined to form an extended and more complete,  
although artifactual, EBV sequence.  
For features, refer to feature tables of V01555 and M35547.

FEATURES

source

1. 184113  
/organism="Human herpesvirus 4"  
/db\_xref="taxon:10376"

misc\_feature

1. 152008  
/note="B95-8 sequences (corresponds to 1-152,008 of  
V01555)"

misc\_feature

152009. 152012  
/note="overlap of B95-8 and Raji sequences at B95-8  
deletion point (corresponds to 152,009-152,012 in V01555,  
and 1-4 in M35547)"

misc\_feature

153013. 163839  
/note="Raji sequences (corresponds to 5-11,831 of M35547)"

misc\_feature

163840. 163843  
/note="overlap of B95-8 and Raji sequences at B95-8  
deletion point (corresponds to 152,009-152,012 of V01555,  
and 11,832-11,835 of M35547)"

misc\_feature

163844. 184113  
/note="B95-8 sequences (corresponds to 152,013-172,282 of  
V01555)"

BASE COUNT 36002 a 55824 c 54622 g 37665 t  
ORIGIN

Query Match 100.0%; Score 20; DB 14; Length 184113;  
Best Local Similarity 100.0%; Pred. No. 12;





CDS	DPQLSAIAQSDIDDVGNLRMTQDKDLFVLLKKKLNLSLPTAERKQFSQVLTTLKA SRLSDKALLKEKFNPAYSMYLAANNIIGQLNQOQLDKMGKVPTEDDGDEY" /codon_start=1 /transl_table=11 /product="Orf4" /protein_id="AAF82344.1" /db_xref="GI:8996045" /translation="MSQSEPDIDROTALEKVLIKHVMKASKYEKTLTAQAKKQASEH CQAQAEIEIQRDAYOQYDGLKDLADLLGLESSORQYQOTLASSFARLOTLE EMFSDFRMEIVSDHIFIRHSESLOIQHLPPSILKKLKPTLAEMQHITVLCGAFES ALEVNEIHFSPESAQAQTLPHILSLPARCILQSRKALYSKLSQFQGHSCERHINY DDSLHVENGSDDQ" /complement(5959..6534) /codon_start=1 /transl_table=11 /product="Orf5" /protein_id="AAF82343.1" /db_xref="GI:8996044" /translation="MTANPGRFSFACQHEILLAIPGYFHAHSGIYDYPALPOVILQFAH NORLRQYQLPAPDSNTLLLPDFIAHWAQLPKVAMGLVLIHPAPLPMWVETSOYAQ LHQRGNDPRLQQTSPSPOTLLAAGAAQLLAGLQPGKVVYTAASVYMKERHPTVNG DAHEKCTLECDRKSPLCRSLRSISGKPH" /complement(6531..7274) /gene="ysaj" /complement(6531..7274) /gene="ysaj" /note="similar to Shigella flexneri MxiJ" /codon_start=1 /transl_table=11 /product="Ysaj" /protein_id="AAF82342.1" /db_xref="GI:8996043" /translation="MNKTKFLVISATWILSGDNGQVLI TELSORQSNVLAFLQPG VSTRRENGKLGSIKVAQSDFTAVDILRLYNLPSKESVEITQAPPGHSLVASPOAE RSLLSLIEQRLEOSLITIPGVNARVHYSPUNLSRVKPEQOVSIVTYAGNEDPK LWMSKTKLFLTNSFAESFQNSVSVVVERPQORQVKEPQFPVAIVLLIGIAAALMT AGAILLWRKQOONKKEAPSDTEQKL" /complement(7278..7580) /gene="ysai" /complement(7278..7580) /gene="ysai" /note="similar to Shigella flexneri MxiI" /codon_start=1 /transl_table=11 /product="Ysai" /protein_id="AAF82341.1" /db_xref="GI:8996042" /translation="MSLTNIPSPISFAEQNIPEMVPGSAIGSEFQKKIANITVAEE REKALINGADTIDVSNPSLLOFQSRVNNYNIAINFATLTHKGASAIIDSVIKAI" /complement(7587..7590) /evidence-not_experimental /codon_start=1 /transl_table=11 /product="Orf6" /protein_id="AAF82341.1" /db_xref="GI:8996042" /translation="MLAEHPVVIKMLNTPNNEESPLNVQTTGVGSAASVDGAPLP TGRKINSYVAGIMLAALTIVIGKLVNNSKARQVKTLETLQSSSPUTVITSGG QSVLVRTQDLDMSTORLREHYQQPVTLLKIAALEEIEIETELDKOLGLLGLDVL PCQPIVRLKGNYPADPNKI IKQTLTDFLGCPKRSLESYNATELIHKAEQGLT" /complement(8920..9426)
CDS	Query Match 92.08; Score 18.4; DB 1; Length 26104; Best Local Similarity 95.08; Pred. No. 78; Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy	1 ggctgtcaccgcttcttgg 20 
Db	3865 GGCTGTACGCGCTTCTTGG 3884
RESULT 5	
AF369954	
LOCUS	26201 bp DNA linear BCT 07-AUG-2001
DEFINITION	Yersinia enterocolitica strain Al27 ysa gene locus, complete sequence.
ACCESSION	AF369954
VERSION	AF369954.1 GI:15088591
KEYWORDS	Yersinia enterocolitica.
SOURCE	Yersinia enterocolitica
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
REFERENCE	Yersinia. 1 to 26201)
AUTHORS	1 (bases 1 to 26201) Foultier,B.G.F., Mueller,S., Purnelle,B., Troisfontaines,P. and Cornelis,G.R.
TITLE	Direct Submision
JOURNAL	Submitted (13-APR-2001) Microbial Pathogenesis Unit, Christian de Duve Institute of Cellular Pathology, Universite Catholique de Louvain, Av Hippocrate, 74, P.Box 74.49, Brussels B-1200, Belgium
FEATURES	Location/Qualifiers
Source	1..26201
	/organism="Yersinia enterocolitica"
	/strain="Al27"
	/serotype="O:8"
	/db_xref="taxon:630"
	/note="biotype: 18"
gene	/complement(1..2376)
CDS	/gene="ysrS"
	/complement(1..2376)
	/gene="ysrS"
	/note="probable two-component sensor"
	/codon_start=1
	/transl_table=11
	/product="YsrS"
	/protein_id="AAK84109.1"
	/db_xref="GI:15088598"
	/translation="MLMKELYKIYIHWALTLLSSTAYAAEAIVFSPQELDYIKTHPV VNYGIFPNSPPIEKENSEIGELTRDYIDILSTVTGIRKPKVFSNDRRESVNLQSG KLSLTSTSLFAOHTGLISSVPITFTWPLTVTRKATRIATPDDLMGCVSTIDYSS LIEWTKQPGVNYKIVSPEETIAEVINROAEAAVLSPTALYVNNKYPOLKISR PHSATSIAAHPEDQILINIINKAIASISAKQAEALAKWIIISDTNIPLPKDYKLT FYVGLATALLCLLFVYRKRKMLVRLGSKNLELSVIAHELKTPGLIGLTACE GLVDRIVTSQRELANVIVHTRELIIDNLDLSLNAKINAGSVTQNPQOLLAELCDT TVKLFISFAETHCTLVQRYQKQFFLPHFDGTLVSOALNNIVSNAIKHDPGMVLI ECSLQVDFKNMFSIEVIDTGTIPSKVLARLSEPFYQGFQKFSRSDTDPKPGTGLGL FYAKKNNHLTGCHLAIVSOPGVGSRYVTIALPAIAHYAENPLPEGLHIIMPTIEISS LSGETIQLDGCLEPYSAEPLPAARGEIDQLDVAGKHQHLHKGDSVILPRP VYASALYLAITDLCNEEQPLESSDSPELHPTTIIITDSRLLVVEDEPILLVQVEL FSSMGFOVDVAVANTOQAYOSWLOHHHTIIVTDCRLDSDGFEIVRLRLKMQDSPEPV LIIICQSASLKTTDAQARAREVGMIDYIILQKPVAREQWQQLIRDYFASKEKEHD"
gene	/complement(2354..3064)
CDS	/gene="ysrK"
	/complement(2354..3064)
	/note="probable two-component response regulator"
	/codon_start=1
	/transl_table=11
	/product="YsrK"
	/protein_id="AAK84107.1"
	/db_xref="GI:15088596"
	/translation="MTQTKTLNIAIIEDEFSRLALMEILKYPYHHKPEPELQILYSD FSIQVGVCSNTPELITLDSNPEIEVLDDYSLVAGETDENEDASLPQGVGLIKRL LQRLPAKTIIVHTAHKSLAVARLAWAGAWGVQKSDSQLELFAISYVARGKKFFPI ELASLSQPTQHTHTLSGRETEVRLMLLNGLKQKETSIVRLNISFKTVSNKTRAFKK LGITSNTDFKYAHEISL"
CDS	/complement(3093..3605)
	/note="Orf2"
	/codon_start=1
	/transl_table=11
	/product="unknown"
	/protein_id="AAK84125.1"
	/db_xref="GI:15088614"
	/translation="MQLQPVHLSPPPVLIIFLPQDNPMKSRFNTHKKSYSYLLIL

CDS	ILSLVSGCTLPDNRNRYNGVIGQNILNSKTIPTENVKITLSLSONKASEKEYMIG KEYSTPMTKTSKTIPIERLQLPNELSSSQPLNTSVRVEKGLIMMSDKLTPLLRSG EKLLTVNDS" complement(3532. .4017) /note="Orf1" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAK84128.1" /db_xref="GI:15088617" /translation="MTNLENLPTSVIVLCSESVIENSKRKLTQOETVVWLSESCLLK MLLDKCSKEVMVAJWEKRGVITVSSYKVLROLRCSFKKAKQDETLIATLPRI SYGTGKALTAESTSVAPQEKAAKKNKVMHOKGLIDAVITCTTLAATCSYFI G"
gene	complement(4193. .4669) /gene="ysaH" /note="similar to Salmonella typhimurium tagB"
CDS	complement(4193. .4669) /gene="ysaH" /note="similar to Salmonella typhimurium tagB" /codon_start=1 /transl_table=11 /product="ysaH" /protein_id="AAK84116.1" /db_xref="GI:15088605" /translation="MKTIIVYLSITFLSPVIAQASCLSSAAIRWDVPEIILEAIIN ESGKPDARINIKNGSDHGMTOINTIDPLKSQGLIKKQALMOPCTNIEAGAVLL SKFKKGGYSWRAGYHSTAHYRDYKSKIMKIVNGSPDFAQRNASLLRSTPTT" complement(4820. .5323) /note="similar to Yersinia enterocoliica strain 8081 Orf3"
misc_feature	complement(5124. .5816) /note="Orf4" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAK84129.1" /db_xref="GI:15088618" /translation="MSQSPEDIRQTALEKVLKHQVMKASRYEKTUTAOKAKQASHE CQQAQEEIQRVIAQYGGDGLQKLLADLLIGLESSOROYQOTLASSEARLOTLL MFSPRMVYEVSDHFIHQHSESLQIOIHLPSLLKQLPTLAEMOHIITVLGGAFES LEVNNEILHSPESPESAAOITLPHILSLPARCTILQSRKALYSKLEQFGHSGERHDNY DSLHVENGSSDQ"
CDS	complement(5774. .6349) /note="Orf5" /codon_start=1 /transl_table=11 /product="unknown" /protein_id="AAK84123.1" /db_xref="GI:15088612" /translation="MTANPGRFSFACQHEILLAPGYFHAHSGHSDYDIPALPQVLOEAAH NQRLROYQLPAPDSNTLLPDFTAEHWAQLPKVAMALGVLHLLHPALPMMVETSOYQA LHQRNGNPNQLOOTPPSQPTLLAAGAAQLLGLQPFCKVYTARASYMFKFRHPTVNG DADKCSPTLECDRRSLPLCRSNLSRISGKPH" complement(6346. .7089) /gene="ysaJ" complement(6346. .7089) /gene="ysaJ" /note="similar to Salmonella typhimurium prgK"
gene	complement(6346. .7089) /gene="ysaJ" /note="similar to Salmonella typhimurium prgK" /codon_start=1 /transl_table=11 /product="ysaJ" /protein_id="AAK84104.1" /db_xref="GI:15088593" /translation="MNKLTKEFLVSAIWLSCGDNQVLLTSLSRQSNVLAIVLQENG VGTRENGKLGNSIKYARSDFTAVDLLRLNPLSOESVEHIQAFPGDSLVAQSPQAE RSRLLSLTQLEQSLTIPGVNARHVHVPNSNRKVEPQOVSLVTVAGNEDP LWMSKILFLTNFAEASFQNVVVVVERPPLQRQVKSEPPVAPVLLIGIAAALMT AAGAILLWRKQONSKEAPSDSTEQL"
gene	complement(7093. .7395) /gene="ysaI" complement(7093. .7395) /gene="ysaI" /note="similar to Salmonella typhimurium prgJ"
CDS	

CDS	1	gcctgtccaccgctttcttg 20	Query Match 92.0%; Score 18.4; DB 1; Length 26201;
			Best Local Similarity 95.0%; Pred. No. 78;
			Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db	3679	GGCTGTGAGCGCTTCTTGG 3698	
RESULT	6		
AC094771/c			
LOCUS			
DEFINITION	AC094771	Rattus norvegicus clone CH230-5P4, *** SEQUENCING IN PROGRESS ***;	91316 bp DNA linear HTG 20-DEC-2001
ACCESSION	AC094771	53 unordered pieces.	
VERSION	AC094771.2	GI:17941551	

## KEYWORDS

HTG: HTGS\_PHASE1.

Norway rat.

## SOURCE

Rattus norvegicus

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

## REFERENCE

1 (bases 1 to 91316)

## AUTHORS

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,  
Alsbrooks, S.L., Amaral, H.C., Are, J.R., Banks, T., Barbara, J.,  
Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J.,  
Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,  
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,  
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,  
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,  
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,  
Denny-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,  
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,  
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,  
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,  
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,  
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,  
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,  
Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C.,  
Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,  
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,  
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,  
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,  
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,  
Loulsegh, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,  
Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,  
Morgan, M., Morris, M., Miner, G., Miner, Z., Mitchell, J., Mohabbat, K.,  
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,  
Ogih, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,  
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,  
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojebokan, I., Rolfe, M.,  
Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N.,  
Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,  
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,  
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,  
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,  
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
Washington, S., Williams, G., Williamson, A., Wlecsyk, R., Wooden, S.,  
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.

## TITLE

JOURNAL

REFERENCE

2 (bases 1 to 91316)

## AUTHORS

Worley, K.C.

## TITLE

Direct Submission

## JOURNAL

Submitted (15-SEP-2001)

Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Dec 20, 2001 this sequence version replaced gi:15624607.

## COMMENT

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GBIO  
Center clone name: CH230-5P4  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329First call to  
findPhrapList  
Consensus quality: 67923 bases at least Q40  
Consensus quality: 73330 bases at least Q30  
Consensus quality: 77558 bases at least Q20  
Estimated insert size: 60296; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-fp estimation  
Quality coverage: 0.9x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 53 contigs, the true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 3344: contig of 3344 bp in length  
3345 3444: gap of unknown length  
3445 5703: contig of 2259 bp in length  
5704 5803: gap of unknown length  
5804 7745: contig of 1942 bp in length  
7746 7845: gap of unknown length  
7846 10509: contig of 2664 bp in length  
10510 10609: gap of unknown length  
10610 12772: contig of 2163 bp in length  
12773 12872: gap of unknown length  
12873 15228: contig of 2356 bp in length  
15229 15228: gap of unknown length  
15230 17685: contig of 2357 bp in length  
17686 17785: gap of unknown length  
17786 20343: contig of 2558 bp in length  
20344 20443: gap of unknown length  
20444 22778: contig of 2335 bp in length  
22779 22878: gap of unknown length  
22879 24815: contig of 1937 bp in length  
24816 24915: gap of unknown length  
24916 26151: contig of 1236 bp in length  
26152 26251: gap of unknown length  
26252 29348: contig of 3097 bp in length  
29349 29448: gap of unknown length  
29449 31178: contig of 1730 bp in length  
31179 31278: gap of unknown length  
31279 32781: contig of 1503 bp in length  
32782 32881: gap of unknown length  
32882 33997: contig of 1116 bp in length  
33998 34097: gap of unknown length  
34098 35448: contig of 1351 bp in length  
35449 35548: gap of unknown length  
35549 37661: contig of 2113 bp in length  
37662 37761: gap of unknown length  
37762 39290: contig of 1529 bp in length  
39291 39390: gap of unknown length  
39391 40541: contig of 1151 bp in length  
40542 40641: gap of unknown length  
40642 42842: contig of 2201 bp in length  
42843 42942: gap of unknown length  
42943 44216: contig of 1274 bp in length  
44217 44316: gap of unknown length  
44317 46145: contig of 1829 bp in length  
46146 46245: gap of unknown length  
46246 48178: contig of 1933 bp in length  
48179 48278: gap of unknown length  
48279 50054: contig of 1776 bp in length  
50055 50154: gap of unknown length  
50155 51248: contig of 1094 bp in length  
51249 51348: gap of unknown length  
51349 52350: contig of 1002 bp in length  
52351 52450: gap of unknown length  
52451 53758: contig of 1308 bp in length  
53759 53858: gap of unknown length  
53859 55679: contig of 1721 bp in length  
55680 55753: gap of unknown length  
55754 57153: contig of 1474 bp in length  
57154 57253: gap of unknown length  
57254 59386: contig of 2133 bp in length  
59387 59486: gap of unknown length  
59487 60654: contig of 1168 bp in length  
60655 60754: gap of unknown length  
60755 61990: contig of 1236 bp in length



KMLYAGVASLGGFGQDFDFENIGDSFF

VRMVALMSAICFAVENSYYTISRYILLDAPLMFFIAAAVYSEKYYEMYPANSLNAYK  
SLATGALGMASSSSWGLFTVWYGLLICINRLWFMIGDLTKSSKIFKFAKLAFLAF  
LLGVPFALYLVFFYIHFQSLTLDGASFPSEFRSLTKNNKIPQNVVADVGTSIIS  
LRLHSTGTVHTVCRRLHSHDHKPPVSESSDMOKEVSCYGYSGFDGDANDDMVVEIDKK  
DGTKRLVLFHTVCRRLHSHDHKPPVSESSDMOKEVSCYGYSGFDGDANDDMVVEIDKK  
NSAPGVAQERVALDTRKRLRHAMTCYLFSEHVKLPANGFEQOEVCASSGRHDLTL  
WYNNNSNELLPDTRKISYKSPASFKISFIESHKMMHINKNLVPHVYSEQTSWPF  
LLRGISYWGNNRNVLLGNAIVWAVTAFIGIFGLVITELFSWLGKPIPLKDSKVV  
NFHVQVHYLLGFVAFHYAPFLMORQMLHLYPAYYFGILALGHALDITIVSYVFRSK  
ROMGYAVIITFLAASYFFKSPFIYGTPTWQELCKQKQWLSGMDYNCNTYFSSLEE  
YKNQTLTKRESQPAATSTVEEITIEGDPSPYEDLMNEDGKKIFKDTSEGNELDPVVKK  
MLEEGANILKVEKRAVLE"

BASE COUNT 746 a 600 c 616 g 914 t  
ORIGIN

Query Match 84.0%; Score 16.8; DB 8; Length 2876;  
Best Local Similarity 90.0%; Pred. No. 5.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggctgtcacgcgtttcttgg 20  
|||||  
Db 2136 GGCTGTACCGCTTTCATCG 2155

RESULT 12  
LOCUS A37576 3375 bp DNA linear PAT 05-MAR-1997  
DEFINITION Sequence 1 from Patent W09404687.  
ACCESSION A37576  
VERSION A37576.1 GI:2294438  
KEYWORDS  
SOURCE baker's yeast.  
ORGANISM Saccharomyces cerevisiae

REFERENCE  
AUTHORS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
TITLE Saccharomycetales; Saccharomycetaceae; Saccharomycetes;  
JOURNAL 1 (bases 1 to 3375)

COMMENT  
Tanner, W., Strahl-Bolsinger, S., Fleer, R. and Fournier, A.  
MODIFIED FUNGAL CELLS AND METHOD FOR PRODUCING RECOMBINANT PRODUCTS  
Patent: WO 9404687-A 1 03-MAR-1994;  
RHONE-POULENC RORER SA (FR)  
Other publication CA 2140894 940303  
Other publication AU 4948293 940315  
Other publication NO 950534 950405  
Other publication ZA 9305915 950413  
Other publication FI 950628 950213  
Other publication DE 4226971 940217  
Other publication JP 85009757 960206.  
Location/Qualifiers  
1..3375  
/organism="Saccharomyces cerevisiae"  
/db\_xref="taxon:4912"  
532..2985  
/codon\_start=1  
/product="PMT1 GENE"  
/protein\_id="CAA02295.1"  
/db\_xref="GI:2294439"  
/translation="MSEETKYKVEODDPVPELDIKUGVVRPHIVTPDSAEIASLTKTM  
VTKELKLVACLAFTAVIRLHGLAMPDSVVEHVFQGFASQYIKGTFFMDVHPPLA  
KMLYAGVASLGGQFGDFENIGDSFTPTTVLMRFFSASLGALTVILMYMTLRYSG  
VRMVALMSAICFAVENSYYTISRYILLDAPLMFFIAAAVYSEKYYEMYPANSLNAYK  
SLATGALGMASSSSWGLFTVWYGLLICINRLWFMIGDLTKSSKIFKFAKLAFLAF  
LLGVPFALYLVFFYIHFQSLTLDGASFPSEFRSLTKNNKIPQNVVADVGTSIIS  
LRLHSTGTVHTVCRRLHSHDHKPPVSESSDMOKEVSCYGYSGFDGDANDDMVVEIDKK  
DGTKRLVLFHTVCRRLHSHDHKPPVSESSDMOKEVSCYGYSGFDGDANDDMVVEIDKK  
NSAPGVAQERVALDTRKRLRHAMTCYLFSEHVKLPANGFEQOEVCASSGRHDLTL  
WYNNNSNELLPDTRKISYKSPASFKISFIESHKMMHINKNLVPHVYSEQTSWPF  
LLRGISYWGNNRNVLLGNAIVWAVTAFIGIFGLVITELFSWLGKPIPLKDSKVV  
NFHVQVHYLLGFVAFHYAPFLMORQMLHLYPAYYFGILALGHALDITIVSYVFRSK  
ROMGYAVIITFLAASYFFKSPFIYGTPTWQELCKQKQWLSGMDYNCNTYFSSLEE  
YKNQTLTKRESQPAATSTVEEITIEGDPSPYEDLMNEDGKKIFKDTSEGNELDPVVKK  
MLEEGANILKVEKRAVLE"

FEATURES  
source  
1..3375  
/organism="Saccharomyces cerevisiae"  
/db\_xref="taxon:4912"  
532..2985  
/codon\_start=1  
/product="PMT1 GENE"  
/protein\_id="CAA02295.1"  
/db\_xref="GI:2294439"  
/translation="MSEETKYKVEODDPVPELDIKUGVVRPHIVTPDSAEIASLTKTM  
VTKELKLVACLAFTAVIRLHGLAMPDSVVEHVFQGFASQYIKGTFFMDVHPPLA  
KMLYAGVASLGGQFGDFENIGDSFTPTTVLMRFFSASLGALTVILMYMTLRYSG  
VRMVALMSAICFAVENSYYTISRYILLDAPLMFFIAAAVYSEKYYEMYPANSLNAYK  
SLATGALGMASSSSWGLFTVWYGLLICINRLWFMIGDLTKSSKIFKFAKLAFLAF  
LLGVPFALYLVFFYIHFQSLTLDGASFPSEFRSLTKNNKIPQNVVADVGTSIIS  
LRLHSTGTVHTVCRRLHSHDHKPPVSESSDMOKEVSCYGYSGFDGDANDDMVVEIDKK  
DGTKRLVLFHTVCRRLHSHDHKPPVSESSDMOKEVSCYGYSGFDGDANDDMVVEIDKK  
NSAPGVAQERVALDTRKRLRHAMTCYLFSEHVKLPANGFEQOEVCASSGRHDLTL  
WYNNNSNELLPDTRKISYKSPASFKISFIESHKMMHINKNLVPHVYSEQTSWPF  
LLRGISYWGNNRNVLLGNAIVWAVTAFIGIFGLVITELFSWLGKPIPLKDSKVV  
NFHVQVHYLLGFVAFHYAPFLMORQMLHLYPAYYFGILALGHALDITIVSYVFRSK  
ROMGYAVIITFLAASYFFKSPFIYGTPTWQELCKQKQWLSGMDYNCNTYFSSLEE  
YKNQTLTKRESQPAATSTVEEITIEGDPSPYEDLMNEDGKKIFKDTSEGNELDPVVKK  
MLEEGANILKVEKRAVLE"

CDS  
1..3375  
/organism="Saccharomyces cerevisiae"  
/db\_xref="taxon:4912"  
532..2985  
/codon\_start=1  
/product="PMT1 GENE"  
/protein\_id="CAA02295.1"  
/db\_xref="GI:2294439"  
/translation="MSEETKYKVEODDPVPELDIKUGVVRPHIVTPDSAEIASLTKTM  
VTKELKLVACLAFTAVIRLHGLAMPDSVVEHVFQGFASQYIKGTFFMDVHPPLA  
KMLYAGVASLGGQFGDFENIGDSFTPTTVLMRFFSASLGALTVILMYMTLRYSG  
VRMVALMSAICFAVENSYYTISRYILLDAPLMFFIAAAVYSEKYYEMYPANSLNAYK  
SLATGALGMASSSSWGLFTVWYGLLICINRLWFMIGDLTKSSKIFKFAKLAFLAF  
LLGVPFALYLVFFYIHFQSLTLDGASFPSEFRSLTKNNKIPQNVVADVGTSIIS  
LRLHSTGTVHTVCRRLHSHDHKPPVSESSDMOKEVSCYGYSGFDGDANDDMVVEIDKK  
DGTKRLVLFHTVCRRLHSHDHKPPVSESSDMOKEVSCYGYSGFDGDANDDMVVEIDKK  
NSAPGVAQERVALDTRKRLRHAMTCYLFSEHVKLPANGFEQOEVCASSGRHDLTL  
WYNNNSNELLPDTRKISYKSPASFKISFIESHKMMHINKNLVPHVYSEQTSWPF  
LLRGISYWGNNRNVLLGNAIVWAVTAFIGIFGLVITELFSWLGKPIPLKDSKVV  
NFHVQVHYLLGFVAFHYAPFLMORQMLHLYPAYYFGILALGHALDITIVSYVFRSK  
ROMGYAVIITFLAASYFFKSPFIYGTPTWQELCKQKQWLSGMDYNCNTYFSSLEE  
YKNQTLTKRESQPAATSTVEEITIEGDPSPYEDLMNEDGKKIFKDTSEGNELDPVVKK  
MLEEGANILKVEKRAVLE"

BASE COUNT 893 a 705 c 705 g 1072 t

## ORIGIN

Query Match 84.0%; Score 16.8; DB 6; Length 3375;  
Best Local Similarity 90.0%; Pred. No. 5.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggctgtcacgcgtttcttgg 20  
|||||  
Db 2304 GGCTGTACCGCTTTCATCG 2323

RESULT 13  
LOCUS 183618 3375 bp DNA linear PAT 10-AUG-1998  
DEFINITION Sequence 1 from patent US 5714377.  
ACCESSION 183618  
VERSION 183618.1 GI:3407148  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 3375)  
AUTHORS Tanner, W., Strahl-Bolsinger, S., Fleer, R. and Fournier, A.  
TITLE Modified fungal cells and method for producing recombinant products  
JOURNAL Patent: US 5714377-A 1 03-FEB-1998;  
FEATURES Location/Qualifiers  
1..3375  
/organism="unknown"  
893 a 705 c 705 g 1072 t  
ORIGIN

Query Match 84.0%; Score 16.8; DB 6; Length 3375;  
Best Local Similarity 90.0%; Pred. No. 5.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggctgtcacgcgtttcttgg 20  
|||||  
Db 2304 GGCTGTACCGCTTTCATCG 2323

RESULT 14  
LOCUS YSCPMT1A 3375 bp DNA linear PLN 13-OCT-1993  
DEFINITION Saccharomyces cerevisiae protein O-glycosylation (PMT1) gene,  
mature peptide.  
ACCESSION L19169 L19700  
VERSION L19169.1 GI:388210  
KEYWORDS dolichyl-phosphate-mannose synthase; membrane glycoprotein; protein  
glycosylation.  
SOURCE Saccharomyces cerevisiae DNA.  
ORGANISM Saccharomyces cerevisiae  
Fukuyama; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
REFERENCE 1 (bases 1 to 3375)  
AUTHORS Strahl-Bolsinger, S., Immervoll, T., Deutzmann, R. and Tanner, W.  
TITLE The gene for a key enzyme of protein O-glycosylation in  
Saccharomyces cerevisiae  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90 (17), 8164-8168 (1993)  
MEDLINE 93376764  
FEATURES Location/Qualifiers  
1..3375  
/organism="Saccharomyces cerevisiae"  
/db\_xref="taxon:4932"  
434..531  
/function="transcription initiation site"  
532..2985  
/evidence=experimental  
/gene="PMT1"  
532..2982  
/gene="PMT1"  
mat\_peptide  
/EC\_number="2.4.1.109"

YSCPMT1A 3375 bp DNA linear PLN 13-OCT-1993  
DEFINITION Saccharomyces cerevisiae protein O-glycosylation (PMT1) gene,  
mature peptide.  
ACCESSION L19169 L19700  
VERSION L19169.1 GI:388210  
KEYWORDS dolichyl-phosphate-mannose synthase; membrane glycoprotein; protein  
glycosylation.  
SOURCE Saccharomyces cerevisiae DNA.  
ORGANISM Saccharomyces cerevisiae  
Fukuyama; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
REFERENCE 1 (bases 1 to 3375)  
AUTHORS Strahl-Bolsinger, S., Immervoll, T., Deutzmann, R. and Tanner, W.  
TITLE The gene for a key enzyme of protein O-glycosylation in  
Saccharomyces cerevisiae  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90 (17), 8164-8168 (1993)  
MEDLINE 93376764  
FEATURES Location/Qualifiers  
1..3375  
/organism="Saccharomyces cerevisiae"  
/db\_xref="taxon:4932"  
434..531  
/function="transcription initiation site"  
532..2985  
/evidence=experimental  
/gene="PMT1"  
532..2982  
/gene="PMT1"  
mat\_peptide  
/EC\_number="2.4.1.109"

YSCPMT1A 3375 bp DNA linear PLN 13-OCT-1993  
DEFINITION Saccharomyces cerevisiae protein O-glycosylation (PMT1) gene,  
mature peptide.  
ACCESSION L19169 L19700  
VERSION L19169.1 GI:388210  
KEYWORDS dolichyl-phosphate-mannose synthase; membrane glycoprotein; protein  
glycosylation.  
SOURCE Saccharomyces cerevisiae DNA.  
ORGANISM Saccharomyces cerevisiae  
Fukuyama; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
REFERENCE 1 (bases 1 to 3375)  
AUTHORS Strahl-Bolsinger, S., Immervoll, T., Deutzmann, R. and Tanner, W.  
TITLE The gene for a key enzyme of protein O-glycosylation in  
Saccharomyces cerevisiae  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90 (17), 8164-8168 (1993)  
MEDLINE 93376764  
FEATURES Location/Qualifiers  
1..3375  
/organism="Saccharomyces cerevisiae"  
/db\_xref="taxon:4932"  
434..531  
/function="transcription initiation site"  
532..2985  
/evidence=experimental  
/gene="PMT1"  
532..2982  
/gene="PMT1"  
mat\_peptide  
/EC\_number="2.4.1.109"



KDHNEVAVGAYATLPPGWFVYVKNLKCCKDTDFAVSMSEPVLDMAADYOVRYLRKE  
HGIWTEEEARPAKELCPKDGGRCLCYLVFNLPNPTFVFOVRAKVDGENHHKSA  
PKMTSSEPLSIRSCCIVPPYFVNDIGAPCTFWEVDVNPAAETNNVTRYVYVDER  
DPPGDTNWTETLTKVYTNANKMTPYVVAASFNLETLPEPRKVRIGDGVIGGYLNPV  
ISKECKDYRYSSKEHSLVNGQHAPLLEEQKAGMNMRLQLETRLDKMRGAVDGRAR  
GDFEDGYWKGYSAGFAQCYKDGNSGIFGDRVTPSLISRLDEQYAGQEDFKOGYVDFK  
HNLAQRTGGGYSAGFAQCYKDGNSGIFGDRVTPSLISRLDEQYAGQEDFKOGYVDFK  
EGASSRVGERSRFEDESKLQOQSLTELTERLTSLKTKGDEIHSTKIYHVYNOOPEGAT  
MSSTGAQLAQLEELIGSTSRSTLRHHYTPGDYLAQDAEEGYNLSRNRSLASALA  
ANREDKSGKWAQLIDLVSPELDTTIARINNYTSSNDRGQGVVEEKYHRSYNEHS  
SR\*

BASE COUNT 3745 a 3419 c 3063 g 3024 t  
ORIGIN

Query Match

Best Local Similarity 84.0%; Score 16.8; DB 3; Length 13251;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggctgtcaccgctttcttgg 20  
Db 2442 GGCAGTCACCACCTTCTTGG 2423

Search completed: May 18, 2002, 14:58:40  
Job time: 8217 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:26:33 ; Search time 250.21 Seconds  
(without alignments)  
19.634 Million cell updates/sec

Title: US-09-623-329-23  
Perfect score: 20  
Sequence: 1 ggcgtgcacgcctttcttg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	16.8	84.0	3375	1	US-08-381-931B-1
2	16.4	82.0	2997	1	US-08-453-862-1
3	16.4	82.0	2997	2	US-08-452-734A-1
4	16.4	82.0	2997	4	US-08-176-401B-1
5	16.4	82.0	2997	5	PCT-US94-14989-1
6	15.8	79.0	1962	2	US-08-967-101-135
7	15.8	79.0	1962	2	US-08-592-541-135
8	15.8	79.0	1962	3	US-09-124-698-135
9	15.8	79.0	1962	4	US-09-127-480-135
10	15.8	79.0	1964	3	US-08-888-077A-16
11	15.8	79.0	1964	4	US-08-496-841C-135
12	15.8	79.0	48974	4	US-08-920-422-17
13	15.4	77.0	1901	1	US-08-153-848-43
14	15.4	77.0	1901	3	US-09-299-843A-43
15	15.4	77.0	1901	4	US-09-088-337B-43
16	15.4	77.0	1901	5	PCT-US93-11153-43
17	15.2	76.0	192	2	US-08-887-365-34
18	15.2	76.0	1235	1	US-08-095-726-13
19	15.2	76.0	1235	1	US-08-095-726-15
20	15.2	76.0	1235	1	US-08-096-623A-13
21	15.2	76.0	1235	1	US-08-096-623A-15
22	15.2	76.0	1291	2	US-08-887-365-35
23	15.2	76.0	2039	4	US-09-276-531-45
24	14.8	74.0	163	5	PCT-US93-06251-37
25	14.8	74.0	834	2	US-08-416-603-5
26	14.8	74.0	1032	1	US-07-792-259-14
27	14.8	74.0	1305	1	US-08-315-671-1

28	14.8	74.0	1306	2	US-08-331-515A-1	Sequence 1, Appli
29	14.8	74.0	1306	4	US-09-168-406A-1	Sequence 1, Appli
30	14.8	74.0	1307	1	US-08-458-023B-5	Sequence 5, Appli
31	14.8	74.0	1307	2	US-08-858-933-1	Sequence 1, Appli
32	14.8	74.0	1384	1	US-07-792-259-16	Sequence 16, Appli
33	14.8	74.0	1620	3	US-08-814-052-12	Sequence 12, Appli
34	14.8	74.0	1620	3	US-08-814-052-13	Sequence 13, Appli
35	14.8	74.0	11832	2	US-08-416-603-1	Sequence 1, Appli
36	14.4	72.0	954	2	US-08-336-198C-2	Sequence 2, Appli
37	14.4	72.0	1426	2	US-08-284-465-2	Sequence 7, Appli
38	14.4	72.0	1426	2	US-08-284-465-7	Sequence 1, Appli
39	14.4	72.0	12847	1	US-08-550-715-1	Sequence 35, Appli
40	14.2	71.0	2946	1	US-08-346-455B-35	Sequence 35, Appli
41	14.2	71.0	2946	3	US-08-977-221-35	Sequence 35, Appli
42	14.2	71.0	2946	5	PCT-US95-06613-35	Sequence 35, Appli
43	14.2	71.0	3251	1	US-08-346-455B-68	Sequence 68, Appli
44	14.2	71.0	3251	3	US-08-977-221-68	Sequence 68, Appli
45	14.2	71.0	3251	5	PCT-US95-06613-68	Sequence 68, Appli

## ALIGNMENTS

RESULT 1  
US-08-381-931B-1  
; Sequence 1, Application US/08381931B  
; Patent No. 5714377  
; GENERAL INFORMATION:  
; APPLICANT: Tanner, Widmar, et al.  
; TITLE OF INVENTION: Modified fungal cells and method for  
; producing recombinant products.  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sears Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/381,931B  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32861  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3375 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHEetical: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Saccharomyces cerevisiae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 532..2985  
; OTHER INFORMATION: /product= "PMT1 gene"

Query Match 84.0% Score 16.8; DB 1; Length 3375;  
Best Local Similarity 90.0%; Pred. No. 12;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ggtgtcaccgctttcttg 20  
|||||  
Db 2304 GGTGTACCGCTTCTTCATCG 2323

RESULT 2  
US-08-453-862-1/c  
; Sequence 1, Application US/08453862  
; Patent No. 5738999  
; GENERAL INFORMATION:  
; APPLICANT: Segerson, Thomas P.  
; APPLICANT: Kinzie, J. Mark  
; APPLICANT: Mulvihill, Eileen R.  
; APPLICANT: Saugstad, Julie A.  
; APPLICANT: Westbrock, Gary L.  
; TITLE OF INVENTION: L-AP4 SENSITIVE GLUTAMATE RECEPTORS  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/453,862  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/176,401  
; FILING DATE: 30-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parmelee, Steven W.  
; REGISTRATION NUMBER: 31,990  
; REFERENCE/DOCKET NUMBER: 13952-18-2  
; TELEPHONE: 206-467-9600  
; TELEFAX: 415-576-0300  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2997 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 133..2877  
US-08-453-862-1

Query Match 82.0%; Score 16.4; DB 1; Length 2997;  
Best Local Similarity 94.4%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 gctgtcaccgctttcttg 19  
|||||  
Db 2210 GCTGTACCGATTCTTG 2193

RESULT 3  
US-08-452-734A-1/c  
; Sequence 1, Application US/08452734A  
; Patent No. 5831047  
; GENERAL INFORMATION:  
; APPLICANT: Segerson, Thomas P.  
; APPLICANT: Kinzie, J. Mark

; APPLICANT: Mulvihill, Eileen R.  
; APPLICANT: Saugstad, Julie A.  
; APPLICANT: Westbrock, Gary L.  
; TITLE OF INVENTION: L-AP4 SENSITIVE GLUTAMATE RECEPTORS  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/452,734A  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/176,401  
; FILING DATE: 30-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parmelee, Steven W.  
; REGISTRATION NUMBER: 31,990  
; REFERENCE/DOCKET NUMBER: 13952-18-1  
; TELEPHONE: 206-467-9600  
; TELEFAX: 415-576-0300  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2997 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 133..2877  
US-08-452-734A-1

Query Match 82.0%; Score 16.4; DB 2; Length 2997;  
Best Local Similarity 94.4%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 gctgtcaccgctttcttg 19  
|||||  
Db 2210 GCTGTACCGATTCTTG 2193

RESULT 4  
US-08-176-401B-1/c  
; Sequence 1, Application US/08176401B  
; Patent No. 6274330  
; GENERAL INFORMATION:  
; APPLICANT: Segerson, Thomas P.  
; APPLICANT: Kinzie, J. Mark  
; APPLICANT: Mulvihill, Eileen R.  
; APPLICANT: Saugstad, Julie A.  
; APPLICANT: Westbrock, Gary L.  
; TITLE OF INVENTION: L-AP4 SENSITIVE GLUTAMATE RECEPTORS  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/176,401B
  FILING DATE: 30-DECEMBER-1993
  CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
    NAME: Parmelee, Steven W.
    REGISTRATION NUMBER: 31,990
    REFERENCE/DOCKET NUMBER: 13952-18-2
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 206-467-9500
    TELEFAX: 415-576-0300
  INFORMATION FOR SEQ ID NO: 1:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 2997 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
    MOLECULE TYPE: cDNA
    FEATURE:
      NAME/KEY: CDS
      LOCATION: 133..2877
      ORIGIN: 1
      PS-08-176-401B-1

```

Query Match 82.0%; Score 16.4; DB 4; Length 2997;  
Best Local Similarity 94.4%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 gctgtcaccgctttcttg 19  
 |||||  
 Db 2210 GCTGTCACCGATTCTTG 2193

```

1  RESULT      5
2  PCT-US94-14989-1/c
3  ; Sequence 1, Application PC/TUS9414989
4  ; GENERAL INFORMATION:
5  ; APPLICANT:
6  ; TITLE OF INVENTION: L-AP4 SENSITIVE GLUTAMATE RECEPTORS
7  ; NUMBER OF SEQUENCES: 3
8  ; COMPUTER READABLE FORM:
9  ; MEDIUM TYPE: Floppy disk
10 ; COMPUTER: IBM PC compatible
11 ; OPERATING SYSTEM: PC-DOS/MS-DOS
12 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
13 ; CURRENT APPLICATION DATA:
14 ; APPLICATION NUMBER: PCT/US94/14989
15 ; FILING DATE:
16 ; CLASSIFICATION:
17 ; PRIOR APPLICATION DATA:
18 ; FILING DATE: 30-DEC-1993
19 ; APPLICATION NUMBER: US 08/176,401
20 ; ATTORNEY/AGENT INFORMATION:
21 ; NAME: Parmelee, Steven W.
22 ; REGISTRATION NUMBER: 31,990
23 ; REFERENCE/DOCKET NUMBER: 13952-18PC
24 ; TELECOMMUNICATION INFORMATION:
25 ; TELEPHONE: (206) 467-9600
26 ; TELEFAX: (415) 543-5043
27 ; INFORMATION FOR SEQ ID NO: 1:
28 ; SEQUENCE CHARACTERISTICS:
29 ; LENGTH: 2997 base pairs
30 ; TYPE: nucleic acid
31 ; STRANDEDNESS: single
32 ; TOPOLOGY: linear
33 ; MOLECULE TYPE: cDNA
34 ; FEATURE:
35 ; NAME/KEY: CDS
36 ; LOCATION: 133..2877

```

PCT-US94-14989-1

Query Match 82.0%; Score 16.4; DB 5; Length 2997;  
Best Local Similarity 94.4%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 gctgtcaccgccttcttg 19  
 |||||  
 Db 2210 GCTGTACCGATTCTTG 2193

```

RESULT 6
US-08-967-101 135/c
: Sequence 135, Application US/08967101
: Patent No. 5840540
: GENERAL INFORMATION:
: APPLICANT: ST. GEORGE-HYSLOP, PETER H
: APPLICANT: ROMMENS, JOHANNA M
: APPLICANT: FRASER, PAUL E
: TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
: TO ALZHEIMER'S DISEASE
: NUMBER OF SEQUENCES: 183
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: TESTA, HURWITZ & THIBEAULT
: STREET: High Street Tower - 125 High Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02110
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/967,101
: FILING DATE: 10-NOV-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/592,541
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Pitcher, Edmund R.
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 248-7000
: TELEFAX: (617) 248-7100
: INFORMATION FOR SEQ ID NO: 135:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1962 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-967-101-135

```

```
Query Match      79.0%; Score 15.8; DB 2; Length 1962;
Best Local Similarity 89.5%; Pred. No. 35;
Matches 17: Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 1 ggctgtcaacgcgtttcttg 19  
 ||||| || |||||  
 Db 1487 GGCTGGCAACGCTTTCTTG 1469

RESULT 7  
US-08-592-541-135/C  
; Sequence 135, Application US/08S02541  
; Patent No. 5986034  
; GENERAL INFORMATION:  
; APPLICANT: ST. GEORGE-HYSLOP, PETER H  
; APPLICANT: ROMMENS, JOHANNA M

APPLICANT: FRASER, PAUL E  
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
TO ALZHEIMER'S DISEASE  
NUMBER OF SEQUENCES: 183  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
STREET: High Street Tower - 125 High Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/592,541  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Pitcher, Edmund R.  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7000  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 135:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1962 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-592-541-135

Query Match 79.0%; Score 15.8; DB 2; Length 1962;  
Best Local Similarity 89.5%; Pred. No. 35;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggcgtgcacgcgtcttcttg 19  
||||| || |||||  
DB 1487 GCGTGGCAACGCTTCTTG 1469

RESULT 8  
US-09-124-698-135/c  
Sequence 135, Application US/09124698  
Patent No. 6117978  
GENERAL INFORMATION:  
APPLICANT: ST. GEORGE-HYSLOP, PETER H  
APPLICANT: ROMMENS, JOHANNA M  
APPLICANT: FRASER, PAUL E  
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
TO ALZHEIMER'S DISEASE  
NUMBER OF SEQUENCES: 183  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
STREET: High Street tower - 125 High Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/124,698  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/592,541

FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Pitcher, Edmund R.  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7000  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 135:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1962 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-09-124-698-135

Query Match 79.0%; Score 15.8; DB 3; Length 1962;  
Best Local Similarity 89.5%; Pred. No. 35;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggcgtgcacgcgtcttcttg 19  
||||| || |||||  
DB 1487 GCGTGGCAACGCTTCTTG 1469

RESULT 9  
US-09-127-480-135/c  
Sequence 135, Application US/09127480  
Patent No. 6194153  
GENERAL INFORMATION:  
APPLICANT: ST. GEORGE-HYSLOP, PETER H  
APPLICANT: ROMMENS, JOHANNA M  
APPLICANT: FRASER, PAUL E  
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
TO ALZHEIMER'S DISEASE  
NUMBER OF SEQUENCES: 183  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
STREET: High Street Tower - 125 High Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/127,480  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/592,541  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Pitcher, Edmund R.  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7000  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 135:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1962 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-09-127-480-135

Query Match 79.0%; Score 15.8; DB 4; Length 1962;  
Best Local Similarity 89.5%; Pred. No. 35;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggctgtcaccgctttcttg 19  
||||| || |||||  
Db 1487 GGCTGGCAACGCTTCTTG 1469

RESULT 10  
US-08-888-077A-16/c  
; Sequence 16, Application US/08888077A  
; Patent No. 6020143  
; GENERAL INFORMATION:  
; APPLICANT: ST. GEORGE-HYSLOP, PETER H  
; APPLICANT: ROMMENS, JOHANNA M  
; APPLICANT: FRASER, PAUL E  
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE AND USES THEREFOR.  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LERNER, DAVID, LITTENBERG, KRUMHOLZ & MENTLIK  
; STREET: 600 SOUTH AVENUE WEST  
; CITY: WESTFIELD  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 07090-1497  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/888,077A  
; FILING DATE: 03-JUL-1997  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/592,541  
; FILING DATE: 26-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PALISI, THOMAS M  
; REGISTRATION NUMBER: 36,629  
; REFERENCE/DOCKET NUMBER: SCHERING 3.0-017 CIP CIP CIP IV  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (908) 654-5000  
; TELEFAX: (908) 654-7866  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1964 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 188..1588  
; FEATURE:  
; NAME/KEY: misc-feature  
; LOCATION: 1..1964  
; OTHER INFORMATION: /note= "mPsi"  
US-08-888-077A-16

Query Match 79.0%; Score 15.8; DB 3; Length 1964;  
Best Local Similarity 89.5%; Pred. No. 35;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggctgtcaccgctttcttg 19  
||||| || |||||  
Db 1489 GGCTGGCAACGCTTCTTG 1471

RESULT 11  
US-08-496-841C-135/c  
; Sequence 135, Application US/08496841C  
; Patent No. 6210919  
; GENERAL INFORMATION:

; APPLICANT: ST. GEORGE-HYSLOP, PETER H  
; APPLICANT: ROMMENS, JOHANNA M  
; APPLICANT: FRASER, PAUL E  
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
; TITLE OF INVENTION: TO ALZHEIMER'S DISEASE  
; NUMBER OF SEQUENCES: 175  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby, PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10022  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/496,841C  
; FILING DATE: 28-Jun-1995  
; CLASSIFICATION: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Paul F. Fehlner, Ph.D.  
; REGISTRATION NUMBER: 35,135  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 527-7700  
; TELEFAX: (212) 753-6237  
; INFORMATION FOR SEQ ID NO: 135:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1964 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; SEQUENCE DESCRIPTION: SEQ ID NO: 135:  
US-08-496-841C-135

Query Match 79.0%; Score 15.8; DB 4; Length 1964;  
Best Local Similarity 89.5%; Pred. No. 35;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggctgtcaccgctttcttg 19  
||||| || |||||  
Db 1489 GGCTGGCAACGCTTCTTG 1471

RESULT 12  
US-08-920-422-17/c  
; Sequence 17, Application US/08920422A  
; Patent No. 6255473  
; GENERAL INFORMATION:  
; APPLICANT: Vitek, Michael P.  
; APPLICANT: Mitsuda No. 6255473iaki  
; APPLICANT: Roses, Allen D.  
; TITLE OF INVENTION: Presenilin-1 Gene Promoter  
; FILE REFERENCE: VITEKPRESENTILIN  
; CURRENT APPLICATION NUMBER: US/08/920,422A  
; CURRENT FILING DATE: 1997-08-29  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 17  
; LENGTH: 48974  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-08-920-422-17

Query Match 79.0%; Score 15.8; DB 4; Length 48974;  
Best Local Similarity 89.5%; Pred. No. 56;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy 1 ggctgtcaccgctttcttg 19
||||| || |||||
Db 47745 GGCTGGCAACGCTTTCTTG 47727

RESULT 13
US-08-153-848-43/c
; Sequence 43, Application US/08153848
; Patent No. 5759804
; GENERAL INFORMATION:
; APPLICANT: Godiska, Ronald
; APPLICANT: Gray, Patrick W.
; APPLICANT: Schweikart, Vicki L.
; TITLE OF INVENTION: No. 5759804e1 Seven Transmembrane Receptors
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/153,848
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/977,452
; FILING DATE: 17-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5759804and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31794
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1901 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 701..1717
US-08-153-848-43

Query Match 77.0%; Score 15.4; DB 1; Length 1901;
Best Local Similarity 94.1%; Pred. No. 55;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ggctgtcaccgctttct 17
|| |||||
Db 160 GGCTGTCCCGCTTTCT 144

RESULT 14
US-09-299-843A-43/c
; Sequence 43, Application US/09299843A
; Patent No. 6107475
; GENERAL INFORMATION:
; APPLICANT: Godiska, Ronald
; APPLICANT: Gray, Patrick W.
; APPLICANT: Schweikart, Vicki L.
```

```
; TITLE OF INVENTION: No. 6107475e1 Seven Transmembrane Receptors
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,843A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/088,337
; FILING DATE: 01-JUN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,848
; FILING DATE: 17-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/977,452
; FILING DATE: 17-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Jill E. Uhl
; REGISTRATION NUMBER: 43,213
; REFERENCE/DOCKET NUMBER: 27866/32059B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX:
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1901 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 701..1717
US-09-299-843A-43

Query Match 77.0%; Score 15.4; DB 3; Length 1901;
Best Local Similarity 94.1%; Pred. No. 55;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ggctgtcaccgctttct 17
|| |||||
Db 160 GGCTGTCCCGCTTTCT 144

RESULT 15
US-09-088-337B-43/c
; Sequence 43, Application US/09088337B
; Patent No. 6348574
; GENERAL INFORMATION:
; APPLICANT: Godiska, Ronald
; APPLICANT: Gray, Patrick W.
; APPLICANT: Schweikart, Vicki L.
; TITLE OF INVENTION: No. 6348574e1 Seven Transmembrane Receptors
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
```

Mon May 20 10:14:42 2002

```

STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/088,337B
  FILING DATE: 01-Jun-1998
  CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/153,848
  FILING DATE: 17-NOV-1993
  APPLICATION NUMBER: US 07/977,452
  FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
  NAME: No. 6348574and, Greta E.
  REGISTRATION NUMBER: 35,302
  REFERENCE/DOCKET NUMBER: 31794
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (312) 474-6300
  TELEFAX: (312) 474-0448
  TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
  LENGTH: 1901 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: single
  TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
  NAME/KEY: CDS
  LOCATION: 701..1717
  SEQUENCE DESCRIPTION: SEQ ID NO: 43:
US-09-088-337B-43

```

```

Query Match      77.0%; Score 15.4; DB 4; Length 1901;
Best Local Similarity 94.1%; Pred. No. 55;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 ggctgtcaccgtttct 17
        |||
Db      160 GGTGTGTCACCGCTTCT 144

```

Search completed: May 18, 2002, 17:26:35  
Job time: 12837 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:22:15 ; Search time 8624.33 Seconds  
(without alignments)  
31.300 Million cell updates/sec

Title: US-09-623-329-23

Perfect score: 20

Sequence: 1 ggctgtcaccgctttcttg 20

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: \*  
1: em\_estba: \*  
2: em\_esthum: \*  
3: em\_estnu: \*  
4: em\_estnu: \*  
5: em\_estov: \*  
6: em\_estpl: \*  
7: em\_estro: \*  
8: em\_htc: \*  
9: gb\_esti: \*  
10: gb\_est2: \*  
11: gb\_htc: \*  
12: gb\_gss: \*  
13: em\_gss\_hum: \*  
14: em\_gss\_inv: \*  
15: em\_gss\_pln: \*  
16: em\_gss\_vrt: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	
1	17.4	87.0	848	12	CNS0490M	AL280831 Tetraodon	
C	2	17.4	87.0	940	10	BF238567	BF238567 601904904
3	17.4	87.0	1023	12	CNS02D33	AL191928 Tetraodon	
C	4	17	85.0	600	10	BI991566	BI991566 4102-61 M
C	5	16.8	84.0	227	9	BB582487	BB582487 BB582487
C	6	16.8	84.0	374	12	B72569	B72569 RPC111-9H24
C	7	16.8	84.0	395	10	BI277318	BI277318 UI-R-CY0-
C	8	16.8	84.0	430	12	AQ486014	AQ486014 RPT-11-2
C	9	16.8	84.0	439	12	AQ402984	AQ402984 HS-5053_A
C	10	16.8	84.0	465	12	FR0012721	AL003971 F.rubripes
C	11	16.8	84.0	477	12	AQ265548	AQ265548 RPC111-69
C	12	16.8	84.0	504	9	AL597693	AL597693 DRF2P313G
C	13	16.8	84.0	509	10	BI016417	BI016417 QVO-ET014
C	14	16.8	84.0	513	12	AQ141819	AQ141819 HS-3154_B
C	15	16.8	84.0	529	10	BF075740	BF075740 225016_MA
C	16	16.8	84.0	530	10	BF080427	BF080427 231307_MA
C	17	16.8	84.0	537	10	BF075749	BF075749 225028_MA

c	18	16.8	84.0	566	12	AQ347694
c	19	16.8	84.0	674	10	BI278306
c	20	16.8	84.0	693	9	AU167516
c	21	16.8	84.0	703	9	AU168922
c	22	16.8	84.0	709	9	AU220380
c	23	16.8	84.0	716	9	AU171445
c	24	16.8	84.0	716	10	BI137935
c	25	16.8	84.0	845	10	BI603235
c	26	16.8	84.0	912	10	BG436680
c	27	16.8	84.0	922	12	CNS0311E
c	28	16.8	84.0	1042	10	BF664086
c	29	16.6	83.0	917	12	CNS02032
c	30	16.4	82.0	284	12	AZ057878
c	31	16.4	82.0	312	9	AA786829
c	32	16.4	82.0	346	10	BE648609
c	33	16.4	82.0	363	12	BI5258
c	34	16.4	82.0	372	12	AZ247379
c	35	16.4	82.0	387	9	AW125199
c	36	16.4	82.0	460	12	AQ317247
c	37	16.4	82.0	472	9	AI162933
c	38	16.4	82.0	485	9	AA536898
c	39	16.4	82.0	494	12	B37086
c	40	16.4	82.0	554	12	FR0036949
c	41	16.4	82.0	665	12	AQ965093
c	42	16.4	82.0	683	12	AG164450
c	43	16.4	82.0	768	12	CNS042M6
c	44	16.4	82.0	891	12	B06527
c	45	16.4	82.0	918	9	AU118969

## ALIGNMENTS

RESULT 1

CNS0490M

LOCUS

DEFINITION

848 bp DNA linear GSS 21-MAY-2000

Tetraodon nigroviridis genome survey sequence T7 end of clone

093J14 of library G from Tetraodon nigroviridis, genomic survey

sequence.

AL280831

AL280831.1

GI:8019132

GSS; genome survey sequence.

KEYWORDS

Tetraodon nigroviridis.

SOURCE

Tetraodon nigroviridis

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;

Tetraodontidae; Tetraodon.

REFERENCE

1 (bases 1 to 848)

Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,

Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and

Weissenbach,J.

TITLE

Characterization and repeat analysis of the compact genome of the

freshwater pufferfish Tetraodon nigroviridis

Unpublished

2 (bases 1 to 848)

Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,

Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,

Saurin,W. and Weissenbach,J.

TITLE

Human gene number estimate provided by genome wide analysis using

Tetraodon nigroviridis DNA sequence

Unpublished

3 (bases 1 to 848)

Genoscope.

Direct Submission

Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

This sequence is a single read and was generated as part of a large

scale clone-end sequencing project of the Tetraodon nigroviridis

genome. For more information, please take a look at

http://www.genoscope.cns.fr/Tetraodon.

FEATURES

source

1. .848

/organism="Tetraodon nigroviridis"



```

/db_xref="taxon:99883"
/clone="093J14"
/clone_lib="G"
/note="Genoscope sequence ID : COBG093DE07LP1-end : T7"
BASE COUNT 170 a 253 c 225 g 198 t 2 others
ORIGIN

Query Match 87.0%; Score 17.4; DB 12; Length 848;
Best Local Similarity 94.7%; Pred. No. 4.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gctgtcacccgtttcttgg 19
|||||
Db 535 GCTGTCAACCGTTTCTTTG 553

RESULT 2
LOCUS BF238567 940 bp mRNA linear EST 14-NOV-2000
DEFINITION BF238567 Homo sapiens cDNA clone IMAGE:4132656 5',
mRNA sequence.
ACCESSION BF238567
VERSION BF238567.1 GI:11152487
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 940)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM1032 row: h column: 01
High quality sequence stop: 207.
Location/Qualifiers
1. 940
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4132656"
/clone_lib="NIH-MGC_54"
/tissue_types="from chronic myelogenous leukemia"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: bone marrow; Vector: pBMR-LIB (Clontech);
Site 1: Sfil (ggcgcctcgcc); Site 2: Sfil (ggcattatggcc
5'); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCCGAGCGCCGACATG-dt(30)Bst-3'
(where B = A, C, or G, and N = A, C, G, or T). Average
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
BASE COUNT 242 a 252 c 298 g 148 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 940;
Best Local Similarity 94.7%; Pred. No. 4.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 gctgtcacccgtttcttgg 20
|||||

```

```

Db 224 GCTGTCAACCGTTTCTTTG 206

RESULT 3
LOCUS CNS02D33 1023 bp DNA linear GSS 13-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
256H01 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL191928
VERSION AL191928.1 GI:7830032
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 1023)
AUTHORS Roest-Crolius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1023)
AUTHORS Roest-Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1023)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
1..1023
Location/Qualifiers
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="256H01"
/clone_lib="G"
/note="Genoscope sequence ID : COAG256CD01SP1-end :
PUC-Ori"
BASE COUNT 178 a 257 c 312 g 255 t 21 others
ORIGIN

Query Match 87.0%; Score 17.4; DB 12; Length 1023;
Best Local Similarity 94.7%; Pred. No. 5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 gctgtcacccgtttcttgg 20
|||||
Db 804 GCTGTCCCGCTTTCTTTG 822

RESULT 4
LOCUS BI991566/c 500 bp mRNA linear EST 20-DEC-2001
DEFINITION BI991566 4102-61 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
mRNA sequence.
ACCESSION BI991566
VERSION BI991566.1 GI:17962590
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

REFERENCE	1 (bases 1 to 600)
AUTHORS	Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W., White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.
TITLE	Gene expression in the developing mouse retina by EST sequencing and microarray analysis
JOURNAL	Nucleic Acids Res. 29 (24), 4983-4993 (2001)
COMMENT	Contact: Klein WH Department of Biochemistry and Molecular Biology University of Texas M.D. Anderson Cancer Center Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA Tel: 713 792 3646 Fax: 713 790 0329.

```

tax: 113 130 125
FEATURES
  source
    Location/Qualifiers
      1..600
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      /clone_lib="Mouse F14.5 retina lambda ZAP II Library"
      /tissue_type="neural retina"
      /dev_stage="embryonic day 14.5 post-fertilization"
      133 a 160 c 171 g 136 t
BASE COUNT
ORIGIN

```

Query Match	85.0%;	Score 17;	PH 10;	Length 600;
Best Local Similarity	100.0%;	Pred. NO. 6.8e+02;		
Matches 17;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ggcgtgcacgcgctttct	17	
Db	305	GGCTGTCAACGGCTTCT	289	
RESULT	5			
BB582487/c				
LOCUS	BB582487	RIKEN full-length enriched,	227 bp	mRNA
DEFINITION	BB582487	adult male colon Mus musculus		
		cDNA clone 9030404K19 5',		mRNA sequence.

RESULT	5
BB582487/C	
LOCUS	
DEFINITION	227 bp mRNA linear EST 30-NOV-2000 BB582487 RIKEN full-length enriched, adult male colon Mus musculus cDNA clone 9030404K19 5', mRNA sequence.
ACCESSION	BB582487
VERSION	BB582487.1 GI:11479031
KEYWORDS	EST.
SOURCE	house mouse.

**REFERENCE**

Aizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T., Carruinci, P. I. (2005). I to 227.

Hanagaki, T., Hayatsu, N., Hirooka, T., Hirozane, T., Hodojama, Y., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kawai, J., Koijima, Y., Komuro, K., Kusakabe, M., Matsuyama, T., Miyazaki, A., Nakamura, M., Nishi, K., Nomura, K., Nunakazi, D., Okazaki, Y., Okido, T., Owa, C., Sakai, C., Sakai, K., Sasakawa, D., Sato, K., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sugabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tanaka, T., Toyota, T., Wachihi, A., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshiki, A., Muramatsu, M. and Hayashizaki, Y.

**TITLE**

YENKIN Mouse ESTs (Aizawa, K. et al. 2000)

TITLE	COMMENT
<p>RNA2 ESTs (Aizawa, K. et al. 2000)</p> <p>Unpublished (2000)</p>	<p>Contract: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suenhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222</p>

Email: genome-rest@sc.riken.go.jp,  
URL: <http://genome-gsc.riken.go.jp/>  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki,  
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermotranscription and thermotranscription of the thermotranscription  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,  
Y. and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation system. *Genome Res.* 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303, 19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

```

FEATURES
  source
    Location/Qualifiers
      1. .227
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="9030404K19"
        /clone.lib="RIKEN full-length enriched, adult male colon"
        /sex="male"
        /tissue.type="colon"
        /dev_stage="adult"
        /map_pos="nrning"

```

*/note="Site\_1": Sall; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genomic Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'-GAGAGAGAAGATCCCAAGAGCTCTTTTTCCTTTTTTNV 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5'-GAGAGAGATTCTCGATTAAATAAATCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLX 1. Cloning sites: 5' end: Sall; 3' end:*

BASE COUNT	55 a	63 c	61 q	48 t
ORIGIN				

```

Query Match      84.0%; Score 16.8; DB 9; Length 227;
Best Local Similarity 90.0%; Pred. No. 6.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  gactgtcacccqcttctcttg 20
          ||| ||| ||| ||| ||| ||| |||
Db      163  GAGTCTTACGCGTCTCTTCG 144

```

RESULT	6	
B72569/c		
LOCUS		
DEFINITION	h72569	374 bp DNA linear
	RPC111-9H24.TV	RPC1-11 Homo sapiens genomic clone
		RPC1-11-9H24, DNA
		sequence.
ACCESSION	B72569	
VERSION	B72569.1	
KEYWORDS	GI:2711720	
SEQUENCE	GSS.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE  
AUTHORS  
Adams, M.D., Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Golden  
K., Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P., and Venter  
J.C.  
1 (bases 1 to 374)  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

TITLE: Use of BAC End Sequences for Sequence-Ready Map Building  
JOURNAL: Unpublished (1997)  
COMMENT: Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

Email: mdadams@tigr.org  
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/ldb/humgen/bac\_end\_search/bac\_end\_search.html  
 Seq primer: T7  
 Class: BAC ends.

#### FEATURES

##### Location/Qualifiers

1..374  
 /organism="Homo sapiens"  
 /db\_xref="GDB:7503263"  
 /db\_xref="taxon:9606"  
 /clone="RPCI-11-9H24"  
 /clone\_lib="RPCI-11"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
 RPC111 Human Male BAC Library"  
 101 a 94 c 104 g 75 t

#### BASE COUNT

ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 374;  
 Best Local Similarity 90.0%; Pred. No. 7.6e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ggctgtcaccgctttcttgg 20  
 Db 234 GTCTGTCACGCTTCTTGG 215

#### RESULT 7

LOCUS B1277318 395 bp mRNA linear EST 19-JUL-2001  
 DEFINITION UI-R-CYO-bxz-e-02-0-UI.s1 UI-R-CYO Rattus norvegicus CDNA clone

ACCESSION B1277318  
 VERSION B1277318.1 GI:14923087  
 KEYWORDS EST.  
 SOURCE Norway rat.

ORGANISM Rattus norvegicus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

REFERENCE 1 (bases 1 to 395)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477

COMMENT Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the non-normalized brown adipose library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)  
 Seq primer: M13 Forward

#### PolyA-res.

##### Location/Qualifiers

1..395  
 /organism="Rattus norvegicus"  
 /strain="Sprague-Dawley"

#### FEATURES

##### source

/db\_xref="taxon:10116"  
 /clone="UI-R-CYO-bxz-e-02-0-UI"  
 /clone\_lib="UI-R-CYO"  
 /dev\_stage="ADULT"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-CYO library is a non-normalized library constructed from rat brown adipose tissue. For a detailed description of the library from which this clone was derived, please visit our web site at [rategen.uiowa.edu](http://rategen.uiowa.edu). The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)  
 TAG\_LIB=UI-R-CYO  
 TAG\_TISSUE=brown adipose  
 TAG\_SEQ=TTGTC"  
 74 a 91 c 98 g 132 t

#### BASE COUNT

ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 395;  
 Best Local Similarity 90.0%; Pred. No. 7.7e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ggctgtcaccgctttcttgg 20  
 Db 352 GACTGTGACCGCTATCTGG 371

#### RESULT 8

LOCUS AQ486014/c 430 bp DNA linear GSS 24-APR-1999  
 DEFINITION RPCI-11-235B1.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-235B1,  
 DNA sequence.

ACCESSION AQ486014  
 VERSION AQ486014.1 GI:4668062  
 KEYWORDS GSS.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 430)  
 AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter  
 ,J.C.

TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
 Map Building

JOURNAL Unpublished (1997)

COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200  
 Fax: 301 838 0208

Email: [hbeetigr.org](mailto:hbeetigr.org)

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: [http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html).  
 Seq primer: SP6

Class: BAC ends.

#### FEATURES

##### source

Location/Qualifiers  
 1..430  
 /organism="Homo sapiens"  
 /db\_xref="GDB:7589880"  
 /db\_xref="taxon:9606"  
 /clone="RPCI-11-235B1"  
 /clone\_lib="RPCI-11"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
 RPC111 Human Male BAC Library"

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics ([inforesgen.com](http://inforesgen.com)). BAC end search page: [http://www.tigr.org/tldb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html)  
Seq primer: SP6  
Class: BAC ends.

## FEATURES

source

Location/Qualifiers

1..477  
/organism="Homo sapiens"  
/db\_xref="GDB:7526480"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-69p9"  
/clone\_lib="RPCI-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPC111 Human Male BAC Library"

BASE COUNT 128 a 130 c 129 g 89 t 1 others

ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 477;  
Best Local Similarity 90.0%; Pred. No. 8e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggctgtcacgcgtttcttgg 20

Db 32 GTCTGTACCGCTTTCTTGG 13

## RESULT 12

AL597693/c

LOCUS

DEFINITION 504 bp mRNA linear EST 14-AUG-2001  
DKF2p313G0515.r1 313 (synonym: hlcc2) Homo sapiens cDNA clone

ACCESSION

VERSION

AL597693

KEYWORDS

EST.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 504)

Bloecker, H., Boecher, M., Brandt, P., Mewes, W., Weil, B. and Wiemann, S.

EST (Bloecker, H., Boecher, M., Brandt, P., Mewes, W., Weil, B. and Wiemann, S.)

Unpublished (1999)

Contact: Bloecker H

MIPS

Am Klopferspitz 18a D-82152 Martinsried, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

sequenced by GBF (National Research Centre for Biotechnology Ltd.,

Braunschweig/Germany) within the cDNA sequencing consortium of the

German Genome Project.

No sl sequence available.

This clone (DKF2p313G0515) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

Location/Qualifiers

1..504

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="DKF2p313G0515"

/clone\_lib="313 (synonym: hlcc2)"

/dev\_stage="adult"

/lab\_host="DH10B"

/note="Vector: ppr1p1Ex2; Site\_1: SfiIA; Site\_2: SfiIB;

cDNA-collection"

BASE COUNT 161 a 83 c 92 g 168 t

ORIGIN

## Query Match

Best Local Similarity 90.0%; Pred. No. 8.2e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggctgtcacgcgtttcttgg 20

Db 330 GGCTGTACCGCTTTCTTGG 311

## RESULT 13

BI016417/c

LOCUS

DEFINITION

QV0-ET0149-150301-519-e06 ET0149 Homo sapiens cDNA, mRNA sequence.

ACCESSION

BI016417

VERSION

BI016417.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 509)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

Goldman, G.H., Carvalhal, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=OV0st2-QV0-ET0149-150301-519-e06t3-2001-03-15&t4=1>

Seq primer: puc 18 forward

High quality sequence start: 47

High quality sequence stop: 125.

Location/Qualifiers

1..509

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="ET0149"

/dev\_stage="Adult"

/note="Organ: lung tumor; Vector: puc18; Site\_1: SmaI;

Site\_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the puc 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT 100 a 142 c 136 g 130 t

ORIGIN

## Query Match

Best Local Similarity 90.0%; Pred. No. 8.2e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggctgtcacgcgtttcttgg 20

Db 471 GGCTGTACCGCTTTCTTGG 452

RESULT 14  
 A0141819/c  
 LOCUS  
 DEFINITION HS\_3154\_B1\_G02\_T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3154 Col=3 Row=N, DNA sequence.  
 ACCESSION A0141819  
 VERSION A0141819.1 GI:3532472  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 513)  
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and Hood,L.  
 Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
 99380589  
 Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3687  
 Email: jwallace@u.washington.edu  
 Sequence Tagged Connector  
 Plate: 3154 row: N column: 3  
 Class: BAC ends  
 High quality sequence stop: 513.

FEATURES  
 source  
 Location/Qualifiers  
 1..513  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="Plate=3154 Col=3 Row=N"  
 /clone\_lib="CIT Approved Human Genomic Sperm Library D"  
 /sex="male"  
 /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"  
 BASE COUNT 155 a 99 c 78 g 180 t 1 others  
 ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 513;  
 Best Local Similarity 90.0%; Pred. No. 8.2e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggctgtcacgcgtttcttgg 20  
 ||||| ||| |||||  
 Db 203 GGCTGTACTCTTCTTGG 184

RESULT 15  
 BF075740  
 LOCUS  
 DEFINITION 225016 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.  
 ACCESSION BF075740  
 VERSION BF075740.1 GI:10869251  
 KEYWORDS EST.  
 SOURCE pig.  
 ORGANISM Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 1 (bases 1 to 529)  
 Fahrrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W. and Keele,J.W.  
 Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine  
 Unpublished (2000)  
 JOURNAL  
 COMMENT Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@mail.marc.usda.gov  
 Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -minmatch 12 options.

PCR Primers  
 FORWARD: AGGAAACAGCTATGACCAT  
 BACKWARD: GTTTCCTCCAGTCACGACG  
 Plate: 42 row: D column: 8  
 Seq primer: ATTTAGGTGACACTATAG.

FEATURES  
 source

Location/Qualifiers  
 1..529  
 /organism="Sus scrofa"  
 /db\_xref="taxon:9823"  
 /clone\_lib="MARC 2P1G"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV SPORT6; Site 1: XbaI; Site 2: XhoI;  
 Library made from pooled tissue from testis, ovary, and endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 114 a 137 c 171 g 106 t 1 others  
 ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 529;  
 Best Local Similarity 90.0%; Pred. No. 8.3e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggctgtcacgcgtttcttgg 20  
 ||||| ||| |||||  
 Db 225 GGCTGTACTCTTCTTGG 244

Search completed: May 18, 2002, 17:22:19  
 Job time: 14690 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:26:32 : Search time 250.21 Seconds  
(without alignments)  
19.634 Million cell updates/sec

Title: US-09-623-329-22

Perfect score: 20

Sequence: 1 caggttcacgtcagctcc 20

Scoring table: IDENTITY NUC

Capopt 10.0 , Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15.8	79.0	847	1	US-08-053-131-184
2	15.8	79.0	847	1	US-08-096-762-184
3	15.8	79.0	847	4	US-09-042-353-47
4	15.8	79.0	847	4	US-08-758-417A-312
5	15.8	79.0	5973	4	US-09-245-041-4
6	15.2	76.0	663	3	US-09-013-067A-1
7	15.2	76.0	756	3	US-09-013-067A-7
8	15.2	76.0	792	1	US-08-403-545-6
9	15.2	76.0	792	4	US-08-404-381-6
10	15.2	76.0	3012	1	US-08-453-742-1
11	15.2	76.0	3012	1	US-08-454-464-1
12	15.2	76.0	3012	1	US-08-453-222-1
13	15.2	76.0	3012	1	US-08-452-802-1
14	15.2	76.0	3390	1	US-08-453-742-26
15	15.2	76.0	3390	1	US-08-454-464-26
16	15.2	76.0	3390	1	US-08-453-222-26
17	15.2	76.0	3390	1	US-08-452-802-26
18	15.2	76.0	3416	1	US-08-453-742-24
19	15.2	76.0	3416	1	US-08-454-464-24
20	15.2	76.0	3416	1	US-08-453-222-24
21	15.2	76.0	3416	1	US-08-452-802-24
22	15.2	76.0	6122	4	US-08-403-545-1
23	15.2	76.0	6122	4	US-08-404-381-1
24	14.4	72.0	1575	1	US-08-375-709-18
25	14.4	72.0	1575	1	US-08-752-929-18
26	14.4	72.0	2068	2	US-08-466-589-1
27	14.4	72.0	2068	2	US-08-700-636-1

28 14.4 72.0 2068 3 US-08-467-574-1 Sequence 1, Appli  
29 14.4 72.0 2068 4 US-09-217-345-1 Sequence 1, Appli  
30 14.4 72.0 2277 1 US-08-496-855A-1 Sequence 1, Appli  
31 14.4 72.0 2374 2 US-08-466-589-5 Sequence 5, Appli  
32 14.4 72.0 2374 2 US-08-700-636-5 Sequence 5, Appli  
33 14.4 72.0 2374 3 US-08-467-574-5 Sequence 5, Appli  
34 14.4 72.0 2374 4 US-09-217-345-5 Sequence 5, Appli  
35 14.4 72.0 3681 1 US-08-571-758-3 Sequence 3, Appli  
36 14.4 72.0 3681 1 US-08-909-984A-3 Sequence 3, Appli  
37 14.4 72.0 3681 1 US-08-909-983-3 Sequence 1, Appli  
38 14.4 72.0 37895 1 US-08-375-709-1 Sequence 1, Appli  
39 14.4 72.0 37895 1 US-08-752-929-1 Sequence 1, Appli  
40 14.4 72.0 37895 3 US-09-090-793-1 Sequence 8, Appli  
41 14.2 71.0 50 4 US-08-957-001B-8 Sequence 8, Appli  
42 14.2 71.0 50 4 US-09-496-301-8 Sequence 8, Appli  
43 14.2 71.0 73 2 US-07-916-098A-47 Sequence 47, Appli  
44 14.2 71.0 77 2 US-07-916-098A-46 Sequence 46, Appli  
45 14.2 71.0 77 2 US-07-916-098A-48 Sequence 48, Appli

## ALIGNMENTS

RESULT 1  
US-08-053-131-184  
: Sequence 184, Application US/08053131  
: Patent No. 5661016  
: GENERAL INFORMATION:  
: APPLICANT: Lonberg, Nils  
: APPLICANT: Kay, Robert M.  
: TITLE OF INVENTION: Transgenic No. 5661016-Human Animals for  
: TITLE OF INVENTION: Producing Heterologous Antibodies  
: NUMBER OF SEQUENCES: 197  
: CORRESPONDENCE ADDRESS:  
: ADDRESS: Townsend and Townsend Kourie and Crew  
: STREET: One Market Plaza, Steuart Tower, Suite 200  
: CITY: San Francisco  
: STATE: California  
: COUNTRY: USA  
: ZIP: 94105  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC Compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patent In Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/053,131  
: FILING DATE: 26-APR-1993  
: CLASSIFICATION: 800  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/990,860  
: FILING DATE: 16-DEC-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/810,279  
: FILING DATE: 17-DEC-1991  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/853,408  
: FILING DATE: 18-MAR-1992  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Smith, William M.  
: REGISTRATION NUMBER: 30,223  
: REFERENCE/DOCKET NUMBER: 14643-9-3  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 415-326-2400  
: TELEFAX: 415-326-2422  
: INFORMATION FOR SEQ ID NO: 184:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 847 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: DNA (genomic)  
: FEATURE:

NAME/KEY: CDS  
LOCATION: join(226...279, 405...700)  
US-08-053-131-184

Query Match 79.0%; Score 15.8; DB 1; Length 847;  
Best Local Similarity 89.5%; Pred. No. 32;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 aggttcacgcctcagctcc 20  
||| || |||||  
Db 235 AGGTCCTCGCTCAGCTCC 253

RESULT 2  
US-08-096-762-184  
: Sequence 184, Application US/08096762  
: Patent No. 5814318  
: GENERAL INFORMATION:  
: APPLICANT: Lonberg, Nils  
: APPLICANT: Kay, Robert M.  
: TITLE OF INVENTION: Transgenic No. 5814318-Human Animals for  
: TITLE OF INVENTION: Producing Heterologous Antibodies  
: NUMBER OF SEQUENCES: 210  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Townsend and Townsend Kourie and Crew  
: STREET: One Market Plaza, Stewart Tower, Suite 200  
: CITY: San Francisco  
: STATE: California  
: COUNTRY: USA  
: ZIP: 94105  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION NUMBER: US/08/096,762  
: FILING DATE: 22-JUL-1993  
: CLASSIFICATION: 800  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/053,131  
: FILING DATE: 26-APR-1993  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/990,860  
: FILING DATE: 16-DEC-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/904,068  
: FILING DATE: 23-JUN-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/853,408  
: FILING DATE: 18-MAR-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/810,279  
: FILING DATE: 17-DEC-1991  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/904,068  
: FILING DATE: 23-JUN-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/904,068  
: FILING DATE: 23-JUN-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/853,408  
: FILING DATE: 18-MAR-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/810,279  
: FILING DATE: 17-DEC-1991  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Smith, William M.  
: REGISTRATION NUMBER: 30,223  
: REFERENCE/DOCKET NUMBER: 14643-9-4  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 415-326-2400  
: TELEFAX: 415-326-2422  
: INFORMATION FOR SEQ ID NO: 184:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 847 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: DNA (genomic)  
: FEATURE:  
: NAME/KEY: CDS  
: LOCATION: join(226...279, 405...700)  
US-08-096-762-184

Query Match 79.0%; Score 15.8; DB 1; Length 847;  
Best Local Similarity 89.5%; Pred. No. 32;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 aggttcacgcctcagctcc 20  
||| || |||||  
Db 235 AGGTCCTCGCTCAGCTCC 253

RESULT 3  
US-09-042-353-47  
: Sequence 47, Application US/09042353  
: Patent No. 6255458  
: GENERAL INFORMATION:  
: APPLICANT: Lonberg, Nils  
: APPLICANT: Kay, Robert M.  
: TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for  
: TITLE OF INVENTION: Producing Heterologous Antibodies  
: NUMBER OF SEQUENCES: 421  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Townsend and Townsend and Crew LLP  
: STREET: Two Embarcadero Center, Eighth Floor  
: CITY: San Francisco  
: STATE: California  
: COUNTRY: USA  
: ZIP: 94111-3834  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.30  
: CURRENT APPLICATION NUMBER: US/09/042,353  
: FILING DATE: 13-MAR-1998  
: CLASSIFICATION: 800  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/810,279  
: FILING DATE: 17-DEC-1991  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/853,408  
: FILING DATE: 18-MAR-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/904,068  
: FILING DATE: 23-JUN-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/990,860  
: FILING DATE: 16-DEC-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/053,131  
: FILING DATE: 26-APR-1993  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/096,762  
: FILING DATE: 22-JUL-1993  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/155,301  
: FILING DATE: 18-NOV-1993  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/161,739  
: FILING DATE: 03-DEC-1993  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/165,699  
: FILING DATE: 10-DEC-1993  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/209,741  
: FILING DATE: 09-MAR-1994  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/352,322  
: FILING DATE: 07-DEC-1994  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/544,404  
: FILING DATE: 10-OCT-1995



PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/728,463  
FILING DATE: 10-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US96/16433  
FILING DATE: 10-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/758,417  
FILING DATE: 02-DEC-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US97/21803  
FILING DATE: 01-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph T.  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 014643-009040US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 847 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(226..280, 406..701)  
OTHER INFORMATION: /product= "vk65.15"  
US-09-042-353-47

Query Match 79.0%; Score 15.8; DB 4; Length 847;  
Best Local Similarity 89.5%; Pred. No. 32;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 aggttcacgcctcagctcc 20  
||| || ||||| |||||  
Db 235 AGGTCCTCGCTCAGCTCC 253

RESULT 4  
US-08-758-417A-312  
Sequence 312, Application US/08758417A  
Patent No. 6300129  
GENERAL INFORMATION:  
APPLICANT: Lonberg, Nils  
Kay, Robert M.  
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for  
Producing Heterologous Antibodies  
NUMBER OF SEQUENCES: 417  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/758.417A  
FILING DATE: 02-Dec-1996  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/728,463  
FILING DATE: 10-OCT-1996  
APPLICATION NUMBER: US 08/544,404  
FILING DATE: 10-OCT-1995

APPLICATION NUMBER: US 08/352,322  
FILING DATE: 07-DEC-1994  
APPLICATION NUMBER: US 08/209,741  
FILING DATE: 09-MAR-1994  
APPLICATION NUMBER: US 08/165,699  
FILING DATE: 10-DEC-1993  
APPLICATION NUMBER: US 08/161,739  
FILING DATE: 03-DEC-1993  
APPLICATION NUMBER: US 08/155,301  
FILING DATE: 18-NOV-1993  
APPLICATION NUMBER: US 08/096,762  
FILING DATE: 22-JUL-1993  
APPLICATION NUMBER: US 08/053,131  
FILING DATE: 26-APR-1993  
APPLICATION NUMBER: US 07/990,860  
FILING DATE: 16-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Serafini, Andrew T.  
REGISTRATION/DOCKET NUMBER: 41,303  
REFERENCE/DOCKET NUMBER: 014643-009030US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 312:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 847 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(226..280, 406..701)  
SEQUENCE DESCRIPTION: SEQ ID NO: 312:  
US-08-758-417A-312

Query Match 79.0%; Score 15.8; DB 4; Length 847;  
Best Local Similarity 89.5%; Pred. No. 32;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 aggttcacgcctcagctcc 20  
||| || ||||| |||||  
Db 235 AGGTCCTCGCTCAGCTCC 253

RESULT 5  
US-09-245-041-4/C  
Sequence 4, Application US/09245041  
Patent No. 6274339  
GENERAL INFORMATION:  
APPLICANT: Moore, K.  
Nagle, D.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT  
OF BODY WEIGHT DISORDERS INCLUDING OBESITY  
FILE REFERENCE: 7853-136  
CURRENT APPLICATION NUMBER: US/09/245,041  
CURRENT FILING DATE: 1999-02-05  
EARLIER APPLICATION NUMBER: 60/093,630  
EARLIER FILING DATE: 1998-07-21  
EARLIER APPLICATION NUMBER: 60/104,978  
EARLIER FILING DATE: 1998-10-20  
NUMBER OF SEQ ID NOS: 131  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4  
LENGTH: 5973  
TYPE: DNA  
ORGANISM: Mus musculus  
US-09-245-041-4

Query Match 79.0%; Score 15.8; DB 4; Length 5973;  
Best Local Similarity 89.5%; Pred. No. 38;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 aggttcacgcctcagctcc 20  
 ||||| ||||| ||

Db 908 AGGTTTCATCTCTCAGCACC 890

RESULT 6

US-09-013-067A-1  
 ; Sequence 1, Application US/09013067A  
 ; Patent No. 6057144  
 ; GENERAL INFORMATION:  
 ; APPLICANT: TOKOYASU, KEN  
 ; APPLICANT: MORI, YUTAKA  
 ; APPLICANT: HAMATSU, SHIOKA  
 ; APPLICANT: HAYASHI, KIYOSHI  
 ; TITLE OF INVENTION: CHITIN DEACETYLASE GENE, VECTOR CONTAINING  
 ; TITLE OF INVENTION: SAID GENE AND TRASFORMANT  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
 ; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY  
 ; CITY: ARLINGTON  
 ; STATE: VA  
 ; COUNTRY: USA  
 ; ZIP: 22202  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/013,067A  
 ; FILING DATE: 01-JAN-1998  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 9-345737  
 ; FILING DATE: 02-DEC-1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: OBLON, NORMAN F.  
 ; REGISTRATION NUMBER: 24,618  
 ; REFERENCE/DOCKET NUMBER: 8361-0002-0  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 703-413-3000  
 ; TELEFAX: 703-413-2220  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 663 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Colletotrichum lindemuthianum  
 ; STRAIN: ATCC 56676  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..663  
 ; OTHER INFORMATION: /note= "METHOD OF DETERMINING THE  
 ; OTHER INFORMATION: CHARACTERISTICS: E"

Query Match 76.0%; Score 15.2; DB 3; Length 663;  
 Best Local Similarity 85.0%; Pred. No. 61;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 caggttcacgcctcagctcc 20  
 ||||| ||||| ||||| ||

Db 78 CACCTTCACCGCTCAGCTCC 97

RESULT 7

US-09-013-067A-7  
 ; Sequence 7, Application US/09013067A  
 ; Patent No. 6057144  
 ; GENERAL INFORMATION:  
 ; APPLICANT: TOKOYASU, KEN  
 ; APPLICANT: MORI, YUTAKA  
 ; APPLICANT: HAMATSU, SHIOKA  
 ; APPLICANT: HAYASHI, KIYOSHI  
 ; TITLE OF INVENTION: CHITIN DEACETYLASE GENE, VECTOR CONTAINING  
 ; TITLE OF INVENTION: SAID GENE AND TRASFORMANT  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
 ; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY  
 ; CITY: ARLINGTON  
 ; STATE: VA  
 ; COUNTRY: USA  
 ; ZIP: 22202  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/013,067A  
 ; FILING DATE: 01-JAN-1998  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 9-345737  
 ; FILING DATE: 02-DEC-1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: OBLON, NORMAN F.  
 ; REGISTRATION NUMBER: 24,618  
 ; REFERENCE/DOCKET NUMBER: 8361-0002-0  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 703-413-3000  
 ; TELEFAX: 703-413-2220  
 ; INFORMATION FOR SEQ ID NO: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 756 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: other nucleic acid  
 ; DESCRIPTION: /desc = "PCR REACTION PRODUCTS"  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Colletotrichum lindemuthianum  
 ; STRAIN: ATCC 56676  
 ; US-09-013-067A-7

Query Match 76.0%; Score 15.2; DB 3; Length 756;  
 Best Local Similarity 85.0%; Pred. No. 62;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 caggttcacgcctcagctcc 20  
 ||||| ||||| ||||| ||

Db 165 CACCTTCACCGCTCAGCTCC 184

RESULT 8

US-08-403-545-6/c  
 ; Sequence 6, Application US/08403545  
 ; Patent No. 5656483  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sokatch, John R.  
 ; APPLICANT: Sykes, Pamela Joy  
 ; APPLICANT: Madhusudhan, K.T.  
 ; TITLE OF INVENTION: Genes Encoding Operon and Promoter for  
 ; TITLE OF INVENTION: Branched Chain Keto Acid Dehydrogenase of Pseudomonas putid  
 ; TITLE OF INVENTION: and Methods

NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Carolyn D. Moon  
ADDRESS: Dunlap, Coddling, Perterson and Lee  
STREET: 9400 N. Broadway, Suite 420  
CITY: Oklahoma City  
STATE: Oklahoma  
COUNTRY: USA  
ZIP: 73114  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette 5.25 inch, 360 Kb Storage  
COMPUTER: IBM AT  
OPERATING SYSTEM: MS-DOS Version 3.3  
SOFTWARE: Professional Write 2.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/403.545  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 7/603/781  
FILING DATE: 23-003-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Carolyn D. Moon  
REGISTRATION NUMBER: 33,022  
REFERENCE/DOCKET NUMBER: 5820.101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: Attorney, (405) 478-5344  
TELEFAX: Attorney, (405) 478-5349  
TELEX:  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 792 Base pairs  
TYPE: Nucleic acid  
STRANDEDNESS: Double-stranded  
TOPOLOGY: Circular  
MOLECULE TYPE: Nucleic acid  
FEATURE:  
NAME/KEY: Control region regulating expression of the bkd  
LOCATION: 1-792  
IDENTIFICATION METHOD: SI nuclease and reverse transcriptase  
PUBLICATION INFORMATION:  
AUTHORS: Madhusudhan, K.T.  
AUTHORS: Huang, G.  
AUTHORS: Burns, Gayle  
AUTHORS: Sokatch, John R.  
TITLE: Transcriptional analysis of the promoter region of  
TITLE: the branched chain keto acid dehydrogenase operon of  
JOURNAL: Journal of Bacteriology  
VOLUME: 172  
PAGES: 5655-5663  
DATE: 1990  
US-08-403-545-6

Query Match 76.0%; Score 15.2; DB 1; Length 792;  
Best Local Similarity 85.0%; Pred. No. 62;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cagggttcagctgcagctcc 20  
||| | ||||| |||||  
Db 29 CATGAGCATCGCTCAGCTCC 10

RESULT 9  
US-08-404-381-6/c  
Sequence 6, Application US/08/04381  
Patent No. 6168945  
GENERAL INFORMATION:  
APPLICANT: Sokatch, John R.

APPLICANT: Sykes, Pamela Joy  
APPLICANT: Madhusudhan, K.T.  
TITLE OF INVENTION: Genes Encoding Operon and Promoter for  
TITLE OF INVENTION: Branched Chain Keto Acid Dehydrogenase of Pseudomonas putid  
TITLE OF INVENTION: and Methods  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Carolyn D. Moon  
ADDRESS: Dunlap, Coddling, Perterson and Lee  
STREET: 9400 N. Broadway, Suite 420  
CITY: Oklahoma City  
STATE: Oklahoma  
COUNTRY: USA  
ZIP: 73114  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette 5.25 inch, 360 Kb Storage  
COMPUTER: IBM AT  
OPERATING SYSTEM: MS-DOS Version 3.3  
SOFTWARE: Professional Write 2.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/404.381  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/603,781  
FILING DATE:  
APPLICATION NUMBER: 07/172,148  
FILING DATE: 23-003-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Carolyn D. Moon  
REGISTRATION NUMBER: 33,022  
REFERENCE/DOCKET NUMBER: 5820.101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: Attorney, (405) 478-5344  
TELEFAX: Attorney, (405) 478-5349  
TELEX:  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 792 Base pairs  
TYPE: Nucleic acid  
STRANDEDNESS: Double-stranded  
TOPOLOGY: Circular  
MOLECULE TYPE: Nucleic acid  
FEATURE:  
NAME/KEY: Control region regulating expression of the bkd  
LOCATION: 1-792  
IDENTIFICATION METHOD: SI nuclease and reverse transcriptase  
IDENTIFICATION METHOD: mapping  
PUBLICATION INFORMATION:  
AUTHORS: Madhusudhan, K.T.  
AUTHORS: Huang, G.  
AUTHORS: Burns, Gayle  
AUTHORS: Sokatch, John R.  
TITLE: Transcriptional analysis of the promoter region of  
TITLE: the branched chain keto acid dehydrogenase operon of  
JOURNAL: Journal of Bacteriology  
VOLUME: 172  
PAGES: 5655-5663  
DATE: 1990  
US-08-404-381-6

Query Match 76.0%; Score 15.2; DB 4; Length 792;  
Best Local Similarity 85.0%; Pred. No. 62;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 cagggttcagctgcagctcc 20  
||| | ||||| |||||  
Db 29 CATGAGCATCGCTCAGCTCC 10

RESULT 10  
US-08-453-742-1/c  
; Sequence 1, Application US/08453742  
; Patent No. 5622839  
; GENERAL INFORMATION:  
; APPLICANT: Moore, Emma E  
; APPLICANT: Sheppard, Paul O  
; APPLICANT: Kuestner, Rolf E  
; TITLE OF INVENTION: Human Calcitonin Receptor  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE and CREW  
; STREET: One Market Plaza, Steuart St. Tower,  
; STREET: Twentieth Floor  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94105-1492  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,742  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/100,887  
; FILING DATE:  
; APPLICATION NUMBER: US 07/954,804  
; FILING DATE: 30-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parmelee, Steven W  
; REGISTRATION NUMBER: 31,990  
; REFERENCE/DOCKET NUMBER: 13952-15-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-467-9600  
; TELEFAX: 415-543-5043  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3012 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; IMMEDIATE SOURCE:  
; CLONE: PHOLLEX  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 52..1476  
US-08-453-742-1

Query Match 76.0%; Score 15.2; DB 1; Length 3012;  
Best Local Similarity 85.0%; Pred. No. 70;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 caggttcacgtcagctcc 20  
Db 1405 CTGGTTCATCTCAGCTCC 1386

RESULT 11  
US-08-454-464-1/c  
; Sequence 1, Application US/08454464  
; Patent No. 5674689  
; GENERAL INFORMATION:  
; APPLICANT: Moore, Emma E  
; APPLICANT: Sheppard, Paul O  
; APPLICANT: Kuestner, Rolf E  
; TITLE OF INVENTION: Human Calcitonin Receptor  
; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE and CREW  
; STREET: One Market Plaza, Steuart St. Tower,  
; STREET: Twentieth Floor  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94105-1492  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/454,464  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/100,887  
; FILING DATE: 02-AUG-1993  
; APPLICATION NUMBER: US 07/954,804  
; FILING DATE: 30-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parmelee, Steven W  
; REGISTRATION NUMBER: 31,990  
; REFERENCE/DOCKET NUMBER: 13952-15-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-467-9600  
; TELEFAX: 415-543-5043  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3012 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; IMMEDIATE SOURCE:  
; CLONE: PHOLLEX  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 52..1476  
US-08-454-464-1

Query Match 76.0%; Score 15.2; DB 1; Length 3012;  
Best Local Similarity 85.0%; Pred. No. 70;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 caggttcacgtcagctcc 20  
Db 1405 CTGGTTCATCTCAGCTCC 1386

RESULT 12  
US-08-453-222-1/c  
; Sequence 1, Application US/08453222  
; Patent No. 5674981  
; GENERAL INFORMATION:  
; APPLICANT: Moore, Emma E  
; APPLICANT: Sheppard, Paul O  
; APPLICANT: Kuestner, Rolf E  
; TITLE OF INVENTION: Human Calcitonin Receptor  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE and CREW  
; STREET: One Market Plaza, Steuart St. Tower,  
; STREET: Twentieth Floor  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94105-1492  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

RESULT 13  
US-08-452-802-1/c  
Sequence 1, Application US/08452802  
Patent No. 5683884  
GENERAL INFORMATION:  
APPLICANT: Moore, Emma E  
APPLICANT: Sheppard, Paul O  
APPLICANT: Kuestner, Rolf E  
TITLE OF INVENTION: Human Calcitonin Receptor  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE and CREW  
STREET: One Market Plaza, Steuart St. Tower,  
STREET: Twentieth Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94105-1492  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/452,802  
FILING DATE: 30-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/100,887  
FILING DATE: 02-AUG-1993

```

RESULT 14
US-08-453-742-26/c
; sequence 26, Application US/08453742
; Patent No. 5622839
; GENERAL INFORMATION:
; APPLICANT: Moore, Emma E
; APPLICANT: Sheppard, Paul O
; APPLICANT: Kuestner, Rolf E
; TITLE OF INVENTION: Human Calcitonin Receptor
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE and CREW,
; STREET: One Market Plaza, Steuart St. Tower,
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105-1492
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453.742
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/08/100,887
; FILING DATE:
; APPLICATION NUMBER: US 07/954,804
; FILING DATE: 30-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 13952-15-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-467-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 26:

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:45:55 ; Search time 1024.22 Seconds  
(without alignments)  
33.526 Million cell updates/sec

Title: US-09-623-329-22

Perfect score: 20

Sequence: 1 cagggttcacgtcagctcc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Genesq\_032802:\*

1:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24:	/SID55/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	AA211672	Oligo specific for
2	17	85.0	3171	AB119819	Drosophila melanog
3	17	85.0	3430	AB116345	Drosophila melanog
C 4	17	85.0	8164	AB119818	Drosophila melanog
C 5	17	85.0	11322	AB116344	Drosophila melanog
6	15.8	79.0	166	ABAF0382	Human foetal liver
7	15.8	79.0	166	ABAF37052	Probe #15518 for g
8	15.8	79.0	166	AAK18619	Human brain expres
9	15.8	79.0	166	AAK44542	Human bone marrow

10	15.8	79.0	166	AA124911	Probe #14844 for g
11	15.8	79.0	166	AA150529	Probe #19215 used
12	15.8	79.0	479	ABA57765	Human foetal liver
13	15.8	79.0	479	ABA27138	Probe #5604 for g
14	15.8	79.0	479	AAK05830	Human brain expres
15	15.8	79.0	479	AAK31459	Human bone marrow
16	15.8	79.0	479	AA115688	Probe #5621 for g
17	15.8	79.0	479	AA137346	Probe #6032 used t
18	15.8	79.0	847	AAQ44225	Human DNA fragment
19	15.8	79.0	847	AAQ78855	Human V-kappa gene
20	15.8	79.0	847	AA137183	DNA fragment vk65
21	15.8	79.0	847	AAV12613	Human DNA fragment
22	15.8	79.0	847	AAV38185	Human DNA fragment
23	15.8	79.0	847	AAV22051	Nucleotide sequenc
C 24	15.8	79.0	1891	AA274428	Human secreted pro
C 25	15.8	79.0	1965	AAA39070	Human secreted pro
C 26	15.8	79.0	2604	AAA48576	CDNA encoding whea
C 27	15.8	79.0	5973	AA291924	Wild type (C57BL/6
C 28	15.8	79.0	8906	AA26715	Human genomic DNA
C 29	15.8	79.0	19795	ABL02606	Drosophila melanog
C 30	15.8	79.0	31132	AA136697	Human musculoskele
C 31	15.8	79.0	31132	AA26716	Human genomic DNA
C 32	15.4	77.0	1788	AA54372	Pseudomonas aerugi
C 33	15.4	77.0	9051	ABL06790	Drosophila melanog
34	15.2	76.0	124	AAK18540	Human brain expres
35	15.2	76.0	124	AAK44458	Human bone marrow
36	15.2	76.0	124	AA150451	Probe #19137 used
C 37	15.2	76.0	164	AA20242	Human gene signatu
C 38	15.2	76.0	251	ABA15462	Human nervous syst
C 39	15.2	76.0	259	ABA11507	Human nervous syst
C 40	15.2	76.0	286	AA237249	Arabidopsis thalia
C 41	15.2	76.0	377	AA293872	Cat flea hindgut a
42	15.2	76.0	447	AAK05748	Human brain expres
43	15.2	76.0	447	AAK31372	Human bone marrow
44	15.2	76.0	447	AAK37265	Probe #5951 used t
45	15.2	76.0	558	AAH10114	Human cDNA clone (

ALIGNMENTS

RESULT 1  
AA211672  
ID AA211672 standard; DNA: 20 BP.

25

XX AA211672;

AC AA211672;

XX 19-NOV-1999 (first entry)

XX Oligo specific for EBV BARF-1 RNA.

DE Epstein Barr Virus; EBV infection; viral; gene transcription: EBER-1;  
KW Epstein Barr early RNA; Epstein Barr nuclear antigen 1; EBNA-1; LMP-1;  
KW latent membrane protein; LMP-2; vIL10; BCRF-1; BARF1; BDLF2; NASBA;  
KW EBV-associated malignancy; primer; ss.

XX Synthetic.

OS Epstein-barr virus.

XX WO9945155-A2.

XX 10-SEP-1999.

XX 01-MAR-1999; 99WO-EP01392.

XX 04-MAR-1998; 98EP-0200655.

XX 14-DEC-1998; 98EP-0204231.

PA (ALKU ) AKZO NOBEL NV.

XX Vervoort MBHJ, Van Den Brule AJC, Middelorp JM;

XX WPI; 1999-551051/46.

XX Identifying Epstein Barr Virus infection -  
 PT Claim 11; Page 20; 50pp; English.  
 PS The invention provides methods for identifying an Epstein Barr Virus  
 XX (EBV) infection, that comprises determining viral gene transcription  
 CC patterns by amplification of specific RNA sequences. The binding sites  
 CC of the oligos suitable for amplification are located in the following  
 CC genes: Epstein Barr early RNA (EBER-1), Epstein Barr nuclear antigen 1  
 CC (EBNA-1), latent membrane protein 1 (LMP-1), LMP-2, and vIL10 (BCRF-1),  
 CC BARF1 and BDLF2. The method comprises (a) amplifying a target sequence  
 CC within one or more RNA(s) transcribed from above gene sequences and the  
 CC (b) detecting the amplified products, determining the transcription  
 CC pattern and identifying the corresponding EBV-associated malignancy. The  
 CC RNA is amplified using a transcription based amplification technique  
 CC such as NASBA. The invention is used to diagnose malignant and  
 CC non-malignant EBV-associated diseases. Sequences AAZ11672-75 represent  
 CC oligos specific for BARF-1 RNA.  
 XX Sequence 20 BP; 3 A; 8 C; 4 G; 5 T; 0 other;  
 SQ

Query Match 100.0%; Score 20; DB 20; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 1.3;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 caggttcacgcgcgcgc 20  
 Db 1 caggttcacgcgcgcgc 20

RESULT 2  
 ABL19819  
 ID ABL19819 standard; DNA; 3171 BP.  
 XX  
 AC ABL19819;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 10930.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 10930; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 XX  
 PT The sequence data for this patent did not form part of the printed  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 10930; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 3171 BP; 955 A; 806 C; 740 G; 670 T; 0 other;  
 Query Match 85.0%; Score 17; DB 23; Length 3171;  
 Best Local Similarity 100.0%; Pred. No. 58;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 caggttcacgcgcgcgc 17  
 Db 748 caggttcacgcgcgcgc 764

RESULT 3  
 ABL16345  
 ID ABL16345 standard; DNA; 3430 BP.  
 XX  
 AC ABL16345;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 508.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 508; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 XX  
 PT The sequence data for this patent did not form part of the printed  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 508; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention



```

Qy      1 caggttcacgtcgcagc 17
      |||||||
Db      1007 caggttcacgtcgcagc 1023

RESULT      4
ABL19818/c
ID      ABL19818 standard; DNA; 8164 BP.
XX
AC      ABL19818;
XX
XX      26-MAR-2002 (first entry)
XX
XX      Drosophila melanogaster genomic polynucleotide SEQ ID NO 10927.
DE
XX      Drosophila; developmental biology; cell signalling; insecticide;
KW
KW      pharmaceutical; gene; ds.
XX
XX      Drosophila melanogaster.
OS
XX      WO200171042-A2.
XX
XX      27-SEP-2001.
XX
XX      23-MAR-2001; 2001WO-US09231.
XX
XX      23-MAR-2000; 2000US-191637P.
PR
XX      11-JUL-2000; 2000US-0614150.
XX
XX      (PEKE ) PE CORP NY.
PA
XX      Venter JC, Adams M, Li PWD, Myers EW;
PI
XX      WPI; 2001-656860/75.
XX
XX      New isolated nucleic acid detection reagent for detecting 1000 or more
PT      genes from Drosophila and for elucidating cell signalling and cell-cell
PT      interactions -
XX
XX      Claim 1: SEQ ID NO 10927; 21pp + Sequence Listing; English.
XX
XX      The invention relates to an isolated nucleic acid detection reagent
CC      capable of detecting 1000 or more genes from Drosophila. The invention is
CC      useful in developmental biology and in elucidating cell signalling and
CC      cell-cell interactions in higher eukaryotes for the development of
CC      insecticides, therapeutics and pharmaceutical drugs. The invention
CC      discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC      sequences (ABL01840-ABL16175) and the encoded proteins
CC      (ABB57737-ABB72072).
CC      The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX
XX      Claim 1: SEQ ID NO 10927; 21pp + Sequence Listing; English.
XX
XX      The invention relates to an isolated nucleic acid detection reagent
CC      capable of detecting 1000 or more genes from Drosophila. The invention is
CC      useful in developmental biology and in elucidating cell signalling and
CC      cell-cell interactions in higher eukaryotes for the development of
CC      insecticides, therapeutics and pharmaceutical drugs. The invention
CC      discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC      sequences (ABL01840-ABL16175) and the encoded proteins
CC      (ABB57737-ABB72072).
CC      The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX
XX      Sequence 8164 BP; 2296 A; 1589 C; 1831 G; 2448 T; 0 other;

Query Match      85.0%; Score 17; DB 23; Length 8164;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 caggttcacgtcgcagc 17
      |||||||
Db      4506 CAGGTTTCATCGCTCAGC 4490

RESULT      5
ABL16344/c
ID      ABL16344 standard; DNA; 11322 BP.
XX
AC      ABL16344;
XX
XX      26-MAR-2002 (first entry)
DT
XX
XX      Drosophila melanogaster genomic polynucleotide SEQ ID NO 505.
DE
XX      Drosophila; developmental biology; cell signalling; insecticide;
KW
KW      pharmaceutical; gene; ds.
XX
XX      Drosophila melanogaster.
OS
XX      WO200171042-A2.
XX
XX      27-SEP-2001.
XX
XX      23-MAR-2001; 2001WO-US09231.
XX
XX      23-MAR-2000; 2000US-191637P.
PR
XX      11-JUL-2000; 2000US-0614150.
XX
XX      (PEKE ) PE CORP NY.
PA
XX      Venter JC, Adams M, Li PWD, Myers EW;
PI
XX      WPI; 2001-656860/75.
XX
XX      New isolated nucleic acid detection reagent for detecting 1000 or more
PT      genes from Drosophila and for elucidating cell signalling and cell-cell
PT      interactions -
XX
XX      Claim 1: SEQ ID NO 505; 21pp + Sequence Listing; English.
XX
XX      The invention relates to an isolated nucleic acid detection reagent
CC      capable of detecting 1000 or more genes from Drosophila. The invention is
CC      useful in developmental biology and in elucidating cell signalling and
CC      cell-cell interactions in higher eukaryotes for the development of
CC      insecticides, therapeutics and pharmaceutical drugs. The invention
CC      discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC      sequences (ABL01840-ABL16175) and the encoded proteins
CC      (ABB57737-ABB72072).
CC      The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX
XX      Sequence 11322 BP; 3216 A; 2237 C; 2499 G; 3370 T; 0 other;

Query Match      85.0%; Score 17; DB 23; Length 11322;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 caggttcacgtcgcagc 17
      |||||||
Db      4506 CAGGTTTCATCGCTCAGC 4490

RESULT      6
ABA70382
ID      ABA70382 standard; DNA; 166 BP.
XX
XX      ABA70382;
AC
XX
XX      01-FEB-2002 (first entry)
DT
XX
XX      Human foetal liver single exon nucleic acid probe #18687.
DE
XX      Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX      Homo sapiens.
OS
XX      WO200157277-A2.
XX
XX      09-AUG-2001.
XX
XX      30-JAN-2001; 2001WO-US00669.
XX

```

```
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
XX Claim 4; SEQ ID NO 18687; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 166 BP; 31 A; 61 C; 38 G; 36 T; 0 other;

Query Match 79.0%; Score 15.8; DB 22; Length 166;
Best Local Similarity 89.5%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 cagggttcagctcagctc 19
DB 103 caggatcattgctcagctc 121
|||| |||| |||| ||||

RESULT 7
ABA37052
ID ABA37052 standard; DNA; 166 BP.
XX
XX ABA37052;
XX
XX 23-JAN-2002 (first entry)
XX
XX Probe #15518 for gene expression analysis in human heart cell sample.
XX
XX Human; gene expression; heart; microarray; vascular system; probe;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease; ss.
XX
XX Homo sapiens.
XX
XX WO200157274-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000666.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
```

---

```
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX
XX Claim 4; SEQ ID NO 15518; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 166 BP; 31 A; 61 C; 38 G; 36 T; 0 other;

Query Match 79.0%; Score 15.8; DB 22; Length 166;
Best Local Similarity 89.5%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 cagggttcagctcagctc 19
DB 103 caggatcattgctcagctc 121
|||| |||| |||| ||||

RESULT 8
AAK18619
ID AAK18619 standard; DNA; 166 BP.
XX
XX AAK18619;
XX
XX 05-NOV-2001 (first entry)
XX
XX Human brain expressed single exon probe SEQ ID NO: 18610.
XX
XX Human; brain expressed exon; gene expression analysis; probe;
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000667.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
```

PS Example 4; SEQ ID NO: 18610; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.

XX Sequence 166 BP; 31 A; 61 C; 38 G; 36 T; 0 other;

Query Match 79.0%; Score 15.8; DB 22; Length 166;

Best Local Similarity 89.5%; Pred. No. 1.8e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 caggttcacgtcagctc 19

||||| ||||| ||||| |||||

Db 103 caggttcacgtcagctc 121

RESULT 9

AAK44542  
ID AAK44542 standard; DNA; 166 BP.

AC AAK44542;

DT 06-NOV-2001 (first entry)

XX Human bone marrow expressed single exon probe SEQ ID NO: 19099.

XX Human; bone marrow expressed exon; gene expression analysis; probe;

KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

KX Homo sapiens.

OS

XX WO200157276-A2.

PN 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00668.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human bone marrow.

XX Example 4; SEQ ID NO: 19099; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
CC the probes of the invention.

XX Sequence 166 BP; 31 A; 61 C; 38 G; 36 T; 0 other;

Query Match

79.0%; Score 15.8; DB 22; Length 166;

Best Local Similarity 89.5%; Pred. No. 1.8e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 caggttcacgtcagctc 19

||||| ||||| ||||| |||||

Db 103 caggttcacgtcagctc 121

RESULT 10

AA124911

ID AA124911 standard; DNA; 166 BP.

XX AA124911;

XX 12-OCT-2001 (first entry)

XX Probe #14844 for gene expression analysis in human cervical cell sample.

XX Probe; human; microarray; gene expression; cervical epithelial cell;

KW cervical cancer; ss.

XX Homo sapiens.

XX WO200157278-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00670.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human cervical epithelial cells.

XX Claim 25; SEQ ID No 14844; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes  
CC (SENPs). The present sequence is one such probe. The SENPs are derived  
CC from human Hela cells. The SENPs can be used to produce a single exon  
CC microarray, which can be used for measuring human gene expression in a  
CC sample derived from human cervical epithelial cells. By measuring gene  
CC expression, the probes are therefore useful in grading and/or staging  
CC of diseases of the cervix, notably cervical cancer.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 166 BP; 31 A; 61 C; 38 G; 36 T; 0 other;

Query Match 79.0%; Score 15.8; DB 22; Length 166;

Best Local Similarity 89.5%; Pred. No. 1.8e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 caggttcacgtcagctc 19

||||| ||||| ||||| |||||

Db 103 caggttcacgtcagctc 121

RESULT 11

AA150529

```
ID  AA150529 standard; DNA; 166 BP.
XX
XX  AA150529;
AC
XX  17-OCT-2001 (first entry)
DT
XX  Probe #19215 used to measure gene expression in human placenta sample.
DE
XX  Probe #19215 used to measure gene expression in human placenta sample.
XX  Probe; microarray; human; placenta; antenatal diagnosis;
KW  genetic disorder; ss.
XX  Homo sapiens.
XX
XX  WO200157272-A2.
PN
XX  09-AUG-2001.
PD
XX
XX  30-JAN-2001; 2001WO-US00663.
PF
XX  04-FEB-2000; 2000US-0180312.
PR
XX  26-MAY-2000; 2000US-0207456.
PR
XX  30-JUN-2000; 2000US-0608408.
PR
XX  03-AUG-2000; 2000US-0632366.
PR
XX  21-SEP-2000; 2000US-0234687.
PR
XX  27-SEP-2000; 2000US-0236359.
PR
XX  04-OCT-2000; 2000GB-0024263.
XX
XX  (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX  Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX  WPI; 2001-488897/53.
XX
XX  Human genome-derived single exon nucleic acid probes useful for
PT  analyzing gene expression in human placenta -
XX
XX  Claim 25; SEQ ID NO 19215; 654pp; English.
PS
XX  The present invention relates to single exon nucleic acid probes (SENPs).
CC  The present sequence is one such probe. The probes are useful for
CC  producing a microarray for predicting, measuring and displaying gene
CC  expression in samples derived from human placenta. The probes are useful
CC  for antenatal diagnosis of human genetic disorders.
XX
XX  Sequence 166 BP; 31 A; 61 C; 38 G; 36 T; 0 other;
SQ
Query Match 79.0%; Score 15.8; DB 22; Length 166;
Best Local Similarity 89.5%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 caggttcacgtcagctc 19
    ||| ||| ||| ||| |||
Db 103 caggatcattgctcagctc 121
RESULT 12
ABA57765
ID  ABA57765 standard; DNA; 479 BP.
XX
XX  ABA57765;
AC
XX  ABA57765;
XX
XX  01-FEB-2002 (first entry)
DT
XX
XX  Human foetal liver single exon nucleic acid probe #6070.
DE
XX
XX  Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
KW
XX  Homo sapiens.
OS
XX  WO200157277-A2.
PN
XX  09-AUG-2001.
PR
```

```
XX  30-JAN-2001; 2001WO-US00669.
XX
XX  04-FEB-2000; 2000US-0180312.
PR
XX  26-MAY-2000; 2000US-0207456.
PR
XX  30-JUN-2000; 2000US-0608408.
PR
XX  03-AUG-2000; 2000US-0632366.
PR
XX  21-SEP-2000; 2000US-0234687.
PR
XX  27-SEP-2000; 2000US-0236359.
PR
XX  04-OCT-2000; 2000GB-0024263.
XX
XX  (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX  Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX  WPI; 2001-483447/52.
XX
XX  Human genome-derived single exon nucleic acid probes useful for
PT  analyzing gene expression in human fetal liver -
XX
XX  Claim 1; SEQ ID NO 6070; 639pp + sequence listing; English.
PS
XX  The invention relates to a single exon nucleic acid probe for
CC  measuring human gene expression in a sample derived from human foetal
CC  liver. The single exon nucleic acid probes may be used for predicting,
CC  measuring and displaying gene expression in samples derived from human
CC  foetal liver. The present sequence is a single exon nucleic acid
CC  probe of the invention.
CC  Note: The sequence data for this patent did not form part of the
CC  printed specification, but was obtained in electronic format directly
CC  from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX  Sequence 479 BP; 92 A; 162 C; 129 G; 96 T; 0 other;
SQ
Query Match 79.0%; Score 15.8; DB 22; Length 479;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 caggttcacgtcagctc 19
    ||| ||| ||| ||| |||
Db 384 caggatcattgctcagctc 402
RESULT 13
ABA27138
ID  ABA27138 standard; DNA; 479 BP.
XX
XX  ABA27138;
AC
XX  ABA27138;
XX
XX  23-JAN-2002 (first entry)
DT
XX
XX  Probe #5604 for gene expression analysis in human heart cell sample.
DE
XX
XX  Human; gene expression; heart; microarray; vascular system; probe;
KW  cardiovascular disease; hypertension; cardiac arrhythmia;
KW  congenital heart disease; ss.
XX
XX  Homo sapiens.
OS
XX  WO200157274-A2.
PN
XX  09-AUG-2001.
PD
XX
XX  30-JAN-2001; 2001WO-US00666.
PF
XX
XX  04-FEB-2000; 2000US-0180312.
PR
XX  26-MAY-2000; 2000US-0207456.
PR
XX  30-JUN-2000; 2000US-0608408.
PR
XX  03-AUG-2000; 2000US-0632366.
PR
XX  21-SEP-2000; 2000US-0234687.
PR
XX  27-SEP-2000; 2000US-0236359.
PR
XX  04-OCT-2000; 2000GB-0024263.
PR
```

XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-48899/53.  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX hearts -  
XX Claim 1; SEQ ID No 5604; 530pp; English.  
XX The present invention relates to single exon nucleic acid probes for  
XX measuring human gene expression in a sample derived from human heart. The  
XX present sequence is one such probe. The probes may be used for  
XX predicting, measuring and displaying gene expression in samples derived  
XX from the human heart via microarrays. By measuring gene expression, the  
XX probes are useful for predicting, diagnosing, grading, staging,  
XX monitoring and prognosing diseases of the human heart and vascular system  
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
XX congenital heart disease.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 479 BP; 92 A; 162 C; 129 G; 96 T; 0 other;  
SQ

Query Match 79.0%; Score 15.8; DB 22; Length 479;  
Best Local Similarity 89.5%; Pred. No. 1.9e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 cagggttcacgtcgcagctc 19  
||||| ||||| ||||| |||||  
DB 384 caggatcattgctcagctc 402

RESULT 14  
AAK05830  
ID AAK05830 standard; DNA; 479 BP.  
XX  
XX AC AAK05830;  
XX  
XX DT 05-NOV-2001 (first entry)  
XX  
XX DE Human brain expressed single exon probe SEQ ID NO: 5821.  
XX  
XX KW Human; brain expressed exon; gene expression analysis; probe;  
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX epilepsy; cancer; ss.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200157275-A2.  
XX  
XX PD 09-AUG-2001.  
XX  
XX PF 30-JAN-2001; 2001WO-US00667.  
XX  
XX PR 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX DR WPI; 2001-483446/52.

PT Single exon nucleic acid probes for analyzing gene expression in human  
PT brains -  
XX  
XX PS Example 4; SEQ ID NO: 5821; 650pp + Sequence Listing; English.  
XX  
XX CC The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX brain. They can be used to measure gene expression in brain cell samples,  
XX which may enable the diagnosis and improved treatment of nervous system  
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
XX epilepsy and cancers. The present sequence is one of the probes of the  
XX invention.  
XX  
XX SQ Sequence 479 BP; 92 A; 162 C; 129 G; 96 T; 0 other;  
SQ

Query Match 79.0%; Score 15.8; DB 22; Length 479;  
Best Local Similarity 89.5%; Pred. No. 1.9e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 cagggttcacgtcgcagctc 19  
||||| ||||| ||||| |||||  
DB 384 caggatcattgctcagctc 402

RESULT 15  
AAK31459  
ID AAK31459 standard; DNA; 479 BP.  
XX  
XX AC AAK31459;  
XX  
XX DT 06-NOV-2001 (first entry)  
XX  
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 6016.  
XX  
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;  
XX microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200157275-A2.  
XX  
XX PD 09-AUG-2001.  
XX  
XX PF 30-JAN-2001; 2001WO-US00668.  
XX  
XX PR 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX DR WPI; 2001-488900/53.  
XX  
XX PT Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human bone marrow -  
XX  
XX PS Example 4; SEQ ID NO: 6016; 658pp + Sequence Listing; English.  
XX  
XX CC The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX bone marrow. They can be used to measure gene expression in bone marrow  
XX samples, which may enable the improved diagnosis and treatment of cancers  
XX such as lymphoma, leukaemia and myeloma. The present sequence is one of  
XX the probes of the invention.  
XX  
XX SQ Sequence 479 BP; 92 A; 162 C; 129 G; 96 T; 0 other;

Query Match 79.0%; Score 15.8; DB 22; Length 479;  
 Best Local Similarity 89.5%; Pred. No. 1.9e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 caggttcacgcgcagctc 19  
 |||||  
 Db 384 caggatcattgctcagctc 402

Search completed: May 18, 2002, 17:45:56  
 Job time: 13745 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:22:12 ; Search time 8624.33 Seconds  
(without alignments)  
31.300 Million cell updates/sec

Title: US-09-623-329-22  
Perfect score: 20  
Sequence: 1 caggttcacgtcagctcc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estab:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17.4	87.0	175	9 AW445312	AW445312 81330 MAR
C 2	17.4	87.0	373	9 AW326193	AW326193 18399 MAR
C 3	17.4	87.0	453	10 BF775922	BF775922 286201 MA
C 4	17.4	87.0	546	10 BE808145	BE808145 213276 MA
C 5	17.4	87.0	601	10 BM363119	BM363119 BS320054A
C 6	17.4	87.0	857	10 BE547342	BE547342 601073792
C 7	17	85.0	682	10 B1366532	B1366532 RES2088.5
C 8	17	85.0	875	12 CINS0088E	AL051554 Drosophila
C 9	16.8	84.0	233	10 BE815603	BE815603 FM3-BN016
C 10	16.8	84.0	389	10 BF481978	BF481978 FMI_19.G0
C 11	16.8	84.0	465	12 A2801300	A2801300 2M0059008
C 12	16.8	84.0	502	10 BF421406	BF421406 FMI_8.E02
C 13	16.8	84.0	503	10 BF481404	BF481404 FMI_18.F0
C 14	16.8	84.0	507	9 AW285846	AW285846 LG1_237_G
C 15	16.8	84.0	520	10 BG103006	BG103006 RH12_35
C 16	16.8	84.0	543	10 BE364157	BE364157 P11_12.C0
C 17	16.8	84.0	553	12 AQ429992	AQ429992 HS_5061_B

18	16.8	84.0	566	6 BE359487	BE359487 DGL_53.F0
19	16.8	84.0	644	10 BG782876	BG782876 SEAUWC002
20	16.8	84.0	670	12 AZ989767	AZ989767 2M0273B15
21	16.8	84.0	785	12 BH052693	BH052693 RPCI-24-2
22	16.8	84.0	833	12 BH532609	BH532609 BOCOM44TR
23	16.8	84.0	1038	12 CINS04E3H	AL286550 Tetraodon
24	16.8	84.0	1123	10 BF180939	BF180939 601807179
25	16.4	82.0	151	9 AI943363	A1943363 fc79c05.Y
26	16.4	82.0	308	12 AQ100556	AQ100556 HS_3084_B
27	16.4	82.0	382	12 AZ647354	AZ647354 IM0513M16
28	16.4	82.0	403	10 BI824698	BI824698 603033858
29	16.4	82.0	502	12 FR0041438	AL128930 Fugu rubr
30	16.4	82.0	521	12 FR0040595	AL128088 Fugu rubr
31	16.4	82.0	531	9 A1180132	A1180132 EST223870
32	16.4	82.0	532	12 FR0036565	AL124074 Fugu rubr
33	16.4	82.0	542	10 BG520926	BG520926 ps02d03.Y
34	16.4	82.0	568	10 BM001149	BM001149 1031093E0
35	16.4	82.0	705	12 AG116713	AG116713 Pan trogl
36	16.4	82.0	726	10 BG468668	BG468668 602510272
37	16.4	82.0	880	10 BG442899	BG442899 GA_Ea001
38	16.4	82.0	1132	10 BG107188	BG107188 602290912
39	16	80.0	532	9 AW778162	AW778162 tk45b08.Y
40	16	80.0	553	12 TA37A01P	AL453007 T. brucei
41	16	80.0	572	10 BI473346	BI473346 fp42g11.Y
42	16	80.0	597	10 BI449728	BI449728 dae74g06
43	16	80.0	678	12 AQ953129	AQ953129 Sheared D
44	15.8	79.0	164	12 BH053902	BH053902 RPCI-24-9
45	15.8	79.0	178	9 AV415958	AV415958 AV415958

## ALIGNMENTS

RESULT 1  
AW445312/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AW445312 81330 MAR LBOV Bos taurus cdna 5', mRNA sequence.  
AW445312  
AW445312.1 GI:6987074  
EST.  
COW.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT

Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 175)  
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,  
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett  
,G.L., Heaton,M.P., Laegreid,W., Rohrer,G.A., Chitko-McKown,C.G.,  
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and  
Keele,J.W.  
Sequence evaluation of four pooled-tissue normalized bovine cdna  
libraries and construction of a gene index for cattle  
Genome Res. 11 (4), 626-630 (2001)  
21180013  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smithemall.marc.usda.gov

Single pass sequencing. Bases called and trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 20  
and -minmatch 12 options.  
PCR Primers  
FORWARD: AGGAACACGATGACCAT  
BACKWARD: GTTTCCTCCAGTCACGCG  
Plate: 44 row: P column: 13  
Seq primer: ATTAGGTGACACTATAG.  
Location/Qualifiers  
1. 175  
/organism="Bos taurus"  
/db\_xref="taxon:9913"





ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 546)  
AUTHORS Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,  
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett,  
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,  
Portea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and  
Keele, J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA  
libraries and construction of a gene index for cattle

JOURNAL Genome Res. 11 (4): 626-630 (2001)  
MEDLINE 21180013  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -minmatch 12 options.

PCR Primers  
FORWARD: AGCAACAGCTATCACCAT  
BACKWARD: GTTTCCTCCAGTCACGCG  
Plate: 67 row: A column: 14  
Seq primer: ATTAGGTGACACTATAG.  
Location/Qualifiers

FEATURES  
source 1..546  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 2BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from testis, thymus,  
semitendinosus muscle, longissimus muscle, pancreas,  
adrenal, and endometrium."  
BASE COUNT 69 a 186 c 135 g 154 t  
ORIGIN  
Query Match 87.0%; Score 17.4; DB 10; Length 546;  
Best Local Similarity 94.7%; Pred. No. 5.4e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 aggttcacgtcgtcagctcc 20  
||||||| |||||||  
Db 178 AGGTTTCATCCCTCAGCTCC 160

RESULT 5  
LOCUS BM363119/c 601 bp mRNA linear EST 10-JAN-2002  
DEFINITION BS320054A10003 Subtracted Lewin Cattle Spleen cDNA clone  
BM363119  
ACCESSION BM363119.1 GI:18106488  
VERSION  
KEYWORDS EST.  
SOURCE  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 601)  
AUTHORS Lewin, H.A., Soares, M.B., Pardini, J., Liu, L. and Larson, J.H.  
TITLE Subtracted Lewin Cattle Spleen ESTs  
JOURNAL Unpublished (2002)  
COMMENT Contact: Lewin, H. A.  
W. M. Keck Center for Comparative and Functional Genomics  
University of Illinois at Urbana-Champaign  
340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL

61801, USA  
Tel: 217 333 5998  
Fax: 217 244 5617  
Email: h-lewin@uiuc.edu  
Funding for Cattle EST sequencing was provided by the USDA National  
Research Initiative, Project No. 98-35205-6644, and a grant from  
the Japanese Ministry of Agriculture Fisheries and Forestry to  
H.A. Lewin and J.E. Womack. Base-calling/Quality scores: PHRED form  
Washington University Genome Center. Vector-trimming: CROSS\_MATCH  
from Washington University Genome Center PHRAP suite. This sequence  
is vector free and at least 200bp in length. REPEAT IN THE SEQUENCE  
LOW\_COMPLEXITY STRAND (+) ELEMENT CT-rich LOCATION [9,131].  
Insert length: 601 Std Error: 0.00  
Plate: BS320054A10 row: D column: 03  
Seq primer: CGCAACAGCTCGAAATTAACCC  
High quality sequence stop: 601.  
Location/Qualifiers

FEATURES  
source 1..601  
/organism="Bos taurus"  
/strain="Angus"  
/db\_xref="taxon:9913"  
/clone="BS320054A10D03"  
/clone\_lib="Subtracted Lewin Cattle Spleen"  
/sex="female"  
/dev\_stage="Adult"  
/note="Organ: Spleen; Vector: pBluescript SK(+/-); Site\_1:  
tcORI; Site\_2: XhoI; This BS cDNA library was generated by  
subtraction of the original non-normalized bovine spleen  
library with 16,800 previously sequenced clones from a  
bovine Placenta cDNA library. The original library was  
constructed as described by Band et al (2000), Genome  
Research 10(9): 1359-1368."  
BASE COUNT 77 a 206 c 150 g 168 t  
ORIGIN  
Query Match 87.0%; Score 17.4; DB 10; Length 601;  
Best Local Similarity 94.7%; Pred. No. 5.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 aggttcacgtcgtcagctcc 20  
||||||| |||||||  
Db 167 AGGTTTCATCCCTCAGCTCC 149

RESULT 6  
LOCUS BE547342/c 857 bp mRNA linear EST 09-AUG-2000  
DEFINITION 601073792F1 NIH\_MGC\_12 Homo sapiens cDNA clone IMAGE:3459886 5',  
BE547342  
ACCESSION BE547342  
VERSION BE547342.1 GI:9775987  
KEYWORDS EST.  
SOURCE Human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 857)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM8453 row: g column: 23  
High quality sequence start: 119  
High quality sequence stop: 219.

```

FEATURES
  source
    Location/Qualifiers
      1. .857
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:345986"
        /clone_lib="NIH_MGC_12"
        /tissue_type="cervical carcinoma cell line"
        /lab_host="DH10B"
        /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
        Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
        Average insert size 1.4 kb. Library prepared by Life
        Technologies."
      231 a 152 c 268 g 205 t 1 others
BASE COUNT
ORIGIN

  Query Match 87.0%; Score 17.4; DB 10; Length 857;
  Best Local Similarity 94.7%; Pred. No. 6.4e+02;
  Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 cagggttcacgctcagc 19
  |||
DB 252 CTGGTTCATCGCTCAGTC 234

RESULT 7
B1366532
LOCUS
DEFINITION
  B1366532
  RE52088.5prime RE Drosophila melanogaster normalized Embryo pFlc-1
  Drosophila melanogaster cDNA clone RE52088 5 similar to Pak;
  FBan0010295 located on: 3R 83E4-83E5;; 05/14/2001, mRNA sequence.
ACCESSION
  B1366532
VERSION
  B1366532.1 GI:15062560
KEYWORDS
  EST.
SOURCE
  fruit fly.
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
  Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 682)
  Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson
  J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George
  R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S.,
  Mungall, C. J., Nunoo, J., Pacle, J., Paragas, V., Park, S.,
  Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Celniker, S. and Rubin
  G. M.
  BDGP/HMI RE Drosophila EST Project
  Unpublished (2001)
  Contact: Stapleton, M.
  BDGP
  Lawrence Berkeley National Lab
  One Cyclotron Rd, Berkeley, CA 94720, USA
  Fax: 510 486 6798
  Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
  hit genomic AF003675; arm:3R [2160953,2306738]
  estimated_cyto:83E4-83F1; 05/14/2001
  Plate: RE520 row: H column: 4
  High quality sequence stop: 547.
  Location/Qualifiers
    1. .682
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
      /clone_lib="RE Drosophila melanogaster normalized Embryo
      pFlc-1"
      /sex="male and female"
      /dev_stage="0-24 hours mixed stage embryonic"
      /lab_host="DH5-alpha Tona"
      /note="Organ: embryo; Vector: pFlc1; Site_1: XhoI; Site_2:
      BamHI; Library was kindly generated by Piero Carninci at
      the RIKEN. The library was normalized and excised using
      Cre recombinase. Plasmid cDNA library."
    178 a 191 c 209 g 103 t 1 others
BASE COUNT

FEATURES
  source
    Location/Qualifiers
      1. .875
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /clone_lib="RPC1-98"
        /clone="HACR161.01"
        /note="end : 17"
      236 a 245 c 176 g 203 t 15 others
BASE COUNT
ORIGIN

  Query Match 85.0%; Score 17; DB 12; Length 875;
  Best Local Similarity 100.0%; Pred. No. 1e+03;
  Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cagggttcacgctcagc 17
  |||
DB 495 CAGGTTTCATCGCTCAGC 511

RESULT 9
B815603/c
LOCUS
DEFINITION
  B815603
  PM3-BN0169-270500-001-c03 BN0169 Homo sapiens cDNA, mRNA sequence.
ACCESSION
  B815603
VERSION
  B815603.1 GI:10247837

```

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 233)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., G'Harre,  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-PM3-BNU169-270  
500-001-c03st3-2000-05-27&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 63  
High quality sequence start: 233.

FEATURES  
source Location/Qualifiers  
1..233  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BN0169"  
/dev\_stage="Adult"  
/note="Organ: breast, normal; Vector: puc18; Site\_1: Sma1;  
Site\_2: Sma1; A mini-library was made by cloning products  
derived from ORFEST PCR (O.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
57 a 52 c 77 g 47 t  
BASE COUNT  
ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 233;  
Best Local Similarity 90.0%; Pred. No. 7.3e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 cagggttcalcgctcagctcc 20  
||||| ||||| ||||| ||

Db 114 CAGGTTTCATCCCTCAGCGCC 95

RESULT 10  
LOCUS BF481978  
DEFINITION FM1\_19\_G04\_b1\_A003 Floral-Induced Meristem 1 (FM1) Sorghum  
propinquum cDNA, mRNA sequence.

ACCESSION BF481978  
VERSION BF481978.1 GI:11552799  
KEYWORDS EST.  
SOURCE Sorghum propinquum.  
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Sorghum.  
1 (bases 1 to 389)  
REFERENCE Cordonnier-Pratt, M.-M., Gingle, A., Sudman, M., Marsala, C. and Pratt  
AUTHORS L.H.

TITLE An EST database from Sorghum: floral-induced meristems  
JOURNAL Unpublished (2000)  
COMMENT Contact: Cordonnier-Pratt MM  
Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmpratt@uga.edu  
Sequences have been trimmed to exclude polyA, vector and regions  
below Phred quality 16. The threshold for highest quality sequence  
is 20.  
Seq primer: JEN REV  
High quality sequence stop: 387  
POLYA-No.

FEATURES  
source Location/Qualifiers  
1..389  
/organism="Sorghum propinquum"  
/db\_xref="taxon:142711"  
/clone\_lib="Floral-Induced Meristem 1 (FM1)"  
/note="Organ: Floral-induced meristems; Vector:  
pBluescript II from Lambda Zap II; Site\_1: XbaI; Site\_2:  
EcoRI; mature plants were placed in a growth chamber for  
15 days with 16 hr darkness and 8 hr light (flowering is  
induced by short-day conditions); 16 days after being  
returned to the greenhouse under natural long days during  
late April/early May, meristems were harvested. The  
library was made from poly-A RNA in the cloning vector  
Lambda Zap II. Clones to be sequenced were prepared by  
mass excision."  
64 a 129 c 140 g 66 t  
BASE COUNT  
ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 389;  
Best Local Similarity 90.0%; Pred. No. 8.9e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 cagggttcalcgctcagctcc 20  
||||| ||||| ||||| |||||

Db 322 CAGGTTTCATCCCTCAGCGCTCC 341

RESULT 11  
LOCUS AZ801300/c  
DEFINITION 2M0059008R Mouse 10kb plasmid UUCCLM library Mus musculus genomic  
clone UUCG2M0059008 R, DNA sequence.

ACCESSION AZ801300  
VERSION AZ801300.1 GI:12953623  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 465)  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Stokes, R., Tingey, A., von Niederhausern, A.  
and Wright, D. Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLCT, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: rdunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0059 row: 0 column: 08

Seq primer: CACACAGGAAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 465.

Location/Qualifiers

## FEATURES

source

1. 465  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0059008"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/notes="Vector: pMD22nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

98 a 109 c 115 g 143 t  
BASE COUNT  
ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 465;

Best Local Similarity 90.0%; Pred. No. 9.6e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 caggttcacgtcagctcc 20

||||| ||||| ||||| |||||

Db 415 CAGGTTTCATGTCAGCTCC 396

RESULT 12

BF421406

LOCUS

DEFINITION FMI\_8\_E02.bl\_A003 Floral-Induced Meristem 1 (FMI) Sorghum

502 bp mRNA linear EST 28-NOV-2000

propinquum cDNA, mRNA sequence.

ACCESSION BF421406

VERSION BF421406.1 GI:11409409

KEYWORDS EST.

SOURCE Sorghum propinquum.

ORGANISM Sorghum propinquum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 502)

AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt

,L.H.

TITLE An EST database from Sorghum: floral-induced meristems

JOURNAL Unpublished (2000)

COMMENT Contact: Cordonnier-Pratt MM

Department of Botany

The University of Georgia

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 542 1805

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions

below Phred quality 16. The threshold for highest quality sequence

is 20.

Seq primer: JEN REV

High quality sequence stop: 497

POLYA-No.

## FEATURES

source

Location/Qualifiers

1. 502

/organism="Sorghum propinquum"

/db\_xref="taxon:132711"

/clone\_lib="Floral-Induced Meristem 1 (FMI)"

/note="Organ: Floral-Induced meristems; Vector;

pBluescript II from Lambda Zap II; Site\_1: XhoI; Site\_2:

ECORI; mature plants were placed in a growth chamber for

15 days with 16 hr darkness and 8 hr light (flowering is

induced by short-day conditions); 16 days after being

returned to the greenhouse under natural long days during

late April/early May, meristems were harvested. The

library was made from poly-A RNA in the cloning vector

Lambda Zap II. Clones to be sequenced were prepared by

mass excision."

104 a 118 c 116 g 163 t 1 others

BASE COUNT

ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 502;

Best Local Similarity 90.0%; Pred. No. 9.9e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 caggttcacgtcagctcc 20

||||| ||||| ||||| |||||

Db 120 CAGGTTTCATGTCAGCTCC 139

RESULT 13

BF481404

LOCUS

DEFINITION

FMI\_18\_F09.bl\_A003 Floral-Induced Meristem 1 (FMI) Sorghum

503 bp mRNA linear EST 05-DEC-2000

propinquum cDNA, mRNA sequence.

ACCESSION BF481404

VERSION BF481404.1 GI:11552225

KEYWORDS EST.

SOURCE Sorghum propinquum.

ORGANISM Sorghum propinquum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 503)

AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt

,L.H.

TITLE An EST database from Sorghum: floral-induced meristems

JOURNAL Unpublished (2000)

COMMENT Contact: Cordonnier-Pratt MM

Department of Botany

The University of Georgia

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 542 1805

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions

below Phred quality 16. The threshold for highest quality sequence

is 20.

Seq primer: JEN REV

High quality sequence stop: 394

POLYA-No.

## FEATURES

source

Location/Qualifiers

1. 503

/organism="Sorghum propinquum"

/db\_xref="taxon:132711"

/clone\_lib="Floral-Induced Meristem 1 (FMI)"

/note="Organ: Floral-Induced meristems; Vector;

pBluescript II from Lambda Zap II; Site\_1: XhoI; Site\_2:

ECORI; mature plants were placed in a growth chamber for

15 days with 16 hr darkness and 8 hr light (flowering is

induced by short-day conditions); 16 days after being

returned to the greenhouse under natural long days during

late April/early May, meristems were harvested. The

library was made from poly-A RNA in the cloning vector  
lambda ZAP II. Clones to be sequenced were prepared by  
mass excision."

BASE COUNT 105 a 118 c 116 g 164 t  
ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 507;  
Best Local Similarity 90.0%; Pred. No. 9.9e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 cagggttcacgctcagctcc 20  
|||||  
Db 120 CAGGTTTCATCCTCGGCTCC 139

## RESULT 14

AW285846 507 bp mRNA linear EST 19-JUL-2000  
LOCUS LG1\_237\_G12.bl\_A002 Light Grown 1 (LGI) Sorghum bicolor cDNA, mRNA  
DEFINITION sequence.

ACCESSION AW285846  
VERSION AW285846.2 GI:6858310  
KEYWORDS EST.  
SOURCE Sorghum.

## ORGANISM

Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 507)  
AUTHORS Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C. and Pratt, L.H.  
TITLE An EST database from Sorghum: light-grown seedlings  
JOURNAL Unpublished (2000)  
COMMENT On Jan 6, 2000 this sequence version replaced gi:6675690.

Contact: Cordonnier-Pratt MM  
Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions  
below Phred quality 16. The threshold for highest quality sequence  
is 20.

Seq primer: JEN REV  
High quality sequence stop: 499  
POLYA-No.

## FEATURES

source  
1..507  
Location/Qualifiers  
/organism="Sorghum bicolor"  
/db\_xref="taxon:4558"  
/clone\_lib="Light Grown 1 (LGI)"  
/note="Organ: 10- to 14-day-old light-grown (greenhouse)  
seedlings; Vector: Lambda Zap; Site: 1; XhoI; Site: 2; EcoRI  
; The library was made from poly-A RNA in the cloning  
vector lambda ZAP II. Clones to be sequenced were  
prepared by mass excision."  
75 a 189 c 144 g 99 t

Query Match 84.0%; Score 16.8; DB 9; Length 507;  
Best Local Similarity 90.0%; Pred. No. 9.9e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 cagggttcacgctcagctcc 20  
|||||  
Db 267 CAAGTTCCTCGCTCAGCTCC 286

## RESULT 15

BG103006 520 bp mRNA linear EST 30-JAN-2001  
LOCUS

## DEFINITION

RHIZ2\_35\_H10.bl\_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA  
sequence.

ACCESSION BG103006  
VERSION BG103006.1 GI:12617839  
KEYWORDS EST.  
SOURCE Sorghum propinquum.

## ORGANISM

Sorghum propinquum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 520)  
AUTHORS Cordonnier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt  
L.H.

TITLE An EST database from Sorghum: Sorghum propinquum rhizomes  
JOURNAL Unpublished (2000)  
COMMENT Contact: Cordonnier-Pratt MM

Department of Botany  
The University of Georgia  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 542 1805  
Email: mmpratt@uga.edu

Seq primer: JEN REV  
High quality sequence stop: 492  
POLYA-No.

## FEATURES

Location/Qualifiers  
1..520  
/organism="Sorghum propinquum"  
/db\_xref="taxon:132711"  
/clone\_lib="Rhizome2 (RHIZ2)"  
/note="Organ: Rhizomes; Vector: pBluescript II from Lambda  
Zap II; Site: 1; XhoI; Site: 2; EcoRI; The library was made  
from poly-A RNA in the cloning vector lambda ZAP II.  
Clones to be sequenced were prepared by mass excision."  
88 a 163 c 168 g 101 t

BASE COUNT  
ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 520;  
Best Local Similarity 90.0%; Pred. No. 1e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 cagggttcacgctcagctcc 20  
|||||  
Db 395 CAGGTTTCATCCTCGGCTCC 414

Search completed: May 18, 2002, 17:22:15  
Job time: 14686 sec



## ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 97;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCCTTTACAACCTAAGGC 20

Db 1 CTCCTTTACAACCTAAGGC 20

## RESULT 2

AF120227 LOCUS Human herpesvirus 4 isolate 2854 nuclear antigen EBNA-1 gene, linear VRL 15-OCT-1999

DEFINITION partial cds.

ACCESSION AF120227.1 GI:4585312

VERSION AF120227

KEYWORDS Human herpesvirus 4

SOURCE Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

ORGANISM Gammaherpesvirinae; Lymphocryptovirus.

REFERENCE 1 (bases 1 to 120)

AUTHORS Mackenzie, J., Gray, D., Pinto-Paes, R., Barrezaeta, L.F.,

Armstrong, A.A., Alexander, F.A., McGeoch, D.J. and Jarrett, R.F.

TITLE Analysis of Epstein-Barr virus (EBV) nuclear antigen 1 subtypes in

EBV-associated lymphomas from Brazil and the United Kingdom

J. Gen. Virol. 80 (Pt 10), 2741-2745 (1999)

JOURNAL 20037840

MEDLINE 2 (bases 1 to 120)

REFERENCE Mackenzie, J., Gray, D., Pinto-Paes, R., Barrezaeta, L.F.M.,

Armstrong, A.A., Alexander, F.A., McGeoch, D.J. and Jarrett, R.F.

AUTHORS Direct Submission

TITLE Submitted (14-JAN-1999) LRF Virus Centre, University of Glasgow,

JOURNAL Bearsden Road, Glasgow G61 1QH, UK

FEATURES

source

1..120 /organism="Human herpesvirus 4"

/isolate="2854"

/db\_xref="taxon:10376"

/map="109408-109527"

/note="from patient 2854 (P-Ala)"

<1..>120

/codon\_start=1

/product="nuclear antigen EBNA-1"

/protein\_id="AA025375.1"

/db\_xref="GI:4585313"

/translation="AIIARSHVETTDGTTWAGVFGVGGSKTSLYNI.RRGTAI."

29 a 29 c 35 q 27 t

## BASE COUNT

## ORIGIN

Query Match 100.0%; Score 20; DB 14; Length 120;

Best Local Similarity 100.0%; Pred. No. 57;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCCTTTACAACCTAAGGC 20

Db 87 CTCCTTTACAACCTAAGGC 106

## RESULT 3

AF120228 LOCUS Human herpesvirus 4 isolate 3340 nuclear antigen EBNA-1 gene, linear VRL 15-OCT-1999

DEFINITION partial cds.

ACCESSION AF120228.1 GI:4585314

VERSION AF120228

KEYWORDS Epstein-Barr virus.

SOURCE Human herpesvirus 4

Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 Gammaherpesvirinae; Lymphocryptovirus.

REFERENCE 1 (bases 1 to 120)

AUTHORS Mackenzie, J., Gray, D., Pinto-Paes, R., Barrezaeta, L.F.,

Armstrong, A.A., Alexander, F.A., McGeoch, D.J. and Jarrett, R.F.

TITLE Analysis of Epstein-Barr virus (EBV) nuclear antigen 1 subtypes in

EBV-associated lymphomas from Brazil and the United Kingdom

J. Gen. Virol. 80 (Pt 10), 2741-2745 (1999)

JOURNAL 20037840

MEDLINE 2 (bases 1 to 120)

REFERENCE Mackenzie, J., Gray, D., Pinto-Paes, R., Barrezaeta, L.F.M.,

Armstrong, A.A., Alexander, F.A., McGeoch, D.J. and Jarrett, R.F.

AUTHORS Direct Submission

TITLE Submitted (14-JAN-1999) LRF Virus Centre, University of Glasgow,

JOURNAL Bearsden Road, Glasgow G61 1QH, UK

FEATURES

source

1..120 /organism="Human herpesvirus 4"

/isolate="3340"

/db\_xref="taxon:10376"

/map="109408-109527"

/note="from patient 3340 (V-Ala)"

<1..>120

/codon\_start=1

/product="nuclear antigen EBNA-1"

/protein\_id="AA025376.1"

/db\_xref="GI:4585315"

/translation="AIIARSHVETTDGTTWAGVFGVGGSKTSLYNI.RRGTAI."

29 a 26 c 36 g 29 t

## BASE COUNT

## ORIGIN

Query Match 100.0%; Score 20; DB 14; Length 120;

Best Local Similarity 100.0%; Pred. No. 57;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCCTTTACAACCTAAGGC 20

Db 87 CTCCTTTACAACCTAAGGC 106

## RESULT 4

## AF120230

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## FEATURES

## source

## 1..120

## /organism="Human herpesvirus 4"

## /isolate="3477"

## /db\_xref="taxon:10376"

## /map="109408-109527"

## /note="from patient 3477 (P-Ala)"

```

CDS
<1..>120
/codon_start=1
/product="nuclear antigen EBNA-1"
/protein_id="AAD25378.1"
/db_xref="GI:4585319"
/translation="ALLARSHVERTTEGTWAGVFYVGSGKTSLYNLRGVAL"
29 a 27 c 36 g 28 t

BASE COUNT
ORIGIN
29 a 27 c 36 g 28 t

Query Match 100.0%; Score 20; DB 14; Length 120;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctttacaacctaagc 20
|||||
Db 87 CTCCTTTTACAACTAAGGC 106

RESULT 5
AF120233
LOCUS
DEFINITION
Human herpesvirus 4 clone 3383-11 nuclear antigen EBNA-1 gene,
partial cds.
AF120233
ACCESSION
VERSION
1
SOURCE
Epstein-Barr virus.
Human herpesvirus 4
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
1 (bases 1 to 120)
Mackenzie, J., Gray, D., Pinto-Paes, R., Barrezaeta, L.F.,
Armstrong, A.A., Alexander, F.A., McGeoch, D.J. and Jarrett, R.F.
Analysis of Epstein-Barr virus (EBV) nuclear antigen 1 subtypes in
EBV-associated lymphomas from Brazil and the United Kingdom
J. Gen. Virol. 80 (Pt 10), 2741-2745 (1999)
20037840
2 (bases 1 to 120)
Mackenzie, J., Gray, D., Pinto-Paes, R., Barrezaeta, L.F.M.,
Armstrong, A.A., Alexander, F.A., McGeoch, D.J. and Jarrett, R.F.
Direct Submission
Submitted (14-JAN-1999) LRF Virus Centre, University of Glasgow,
Bearsden Road, Glasgow G61 1QH, UK
1..120
/organism="Human herpesvirus 4"
/db_xref="taxon:10376"
/map="109408-109527"
/clone="3370-30"
<1..>120
/codon_start=1
/product="nuclear antigen EBNA-1"
/protein_id="AAD25381.1"
/db_xref="GI:4585325"
/translation="ALLARSHVERTTEGTWAGVFYVGSGKTSLYNLRGVAL"
29 a 26 c 36 g 29 t

BASE COUNT
ORIGIN
29 a 26 c 36 g 29 t

Query Match 100.0%; Score 20; DB 14; Length 120;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ctccctttacaacctaagc 20
|||||
Db 87 CTCCTTTTACAACTAAGGC 106

RESULT 7
HHV4EBTHR
LOCUS
DEFINITION
Human herpesvirus type 4 EBNA-1 gene (Ala to Thr variation).
ACCESSION
VERSION
X98164
KEYWORDS
EBNA-1 gene.
SOURCE
Epstein-Barr virus.
Human herpesvirus 4
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
1 (bases 1 to 187)
Venkatesh, H., Raj, A., Guitierrez, M.I., Judde, J.G., Spangler, G.,
Variation in the sequence of Epstein Barr virus nuclear antigen 1
in normal peripheral blood lymphocytes and in Burkitt's lymphomas
Oncogene 13 (1), 177-181 (1996)
96292240
2 (bases 1 to 187)
Spangler, G.L.
Direct Submission
Submitted (24-MAY-1996) G.L. Spangler, NCI/NIH, 9000 Rockville
Pike, Building 10 Room 13C206, Bethesda, MD 20892, USA
Related sequence: V01555.
Location/Qualifiers
1..187
/organism="Human herpesvirus 4"

```



```

/db_xref="taxon:10376"
/cell_line="Burkitt's lymphoma"
1..187
/gene="EBNA-1"
<1..>187
/gene="EBNA-1"
/codon_start=1
/protein_id="CAA66845.1"
/db_xref="GI:1359710"
/db_xref="SPTREMBL:Q69543"
/translation="GKHGGGSGNPKFENIAEGLRLLARSHVERTTDEGTWAGVVF
YGSKSTSLYNLRIGLA"
64
/gene="EBNA-1"
/replace="changes codon from Ala to Thr"
176
/gene="EBNA-1"
/replace="C"
175
/gene="EBNA-1"
/replace="C"
53 a 38 c 54 g 42 t
BASE COUNT
ORIGIN

Query Match 100.0%; Score 20; DB 14; Length 187;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ctccctttacaacctaaaggc 20
|||||
Db 150 CTCCTTTTACAACCTAAGGC 169

RESULT 8
AF192740
LOCUS
DEFINITION Human herpesvirus 4 EBNA-1 protein variant P-ala' (EBNA-1) gene,
partial cds.
ACCESSION AF192740
VERSION AF192740.1 GI:6683820
KEYWORDS
SOURCE Epstein-Barr virus.
ORGANISM Human herpesvirus 4
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE 1 (bases 1 to 288)
AUTHORS Fassone,L., Bhatia,K., Gutierrez,M., Capello,D., Gloghini,A.,
Dolcetti,R., Vivenza,D., Ascoli,V., Lo Coco,F., Pagani,L.,
Dotti,G., Rambaldi,A., Raphael,M., Tirelli,U., Saglio,G.,
Magrath,I.T., Carbone,A. and Gaidano,G.
Molecular profile of Epstein Barr virus infection in HHV-8 positive
primary effusion lymphoma
Unpublished
2 (bases 1 to 288)
Fassone,L., Bhatia,K., Gutierrez,M., Capello,D., Gloghini,A.,
Dolcetti,R., Vivenza,D., Ascoli,V., Lo Coco,F., Pagani,L.,
Dotti,G., Rambaldi,A., Raphael,M., Tirelli,U., Saglio,G.,
Magrath,I.T., Carbone,A. and Gaidano,G.
Direct Submission
Submitted (07-OCT-1999) Medical Sciences, Division of Internal
Medicine, A. Avogadro University of Eastern Piedmont, Via Solaroli
17, Novara 28100, Italy
FEATURES
source
Location/Qualifiers
1..288
/organism="Human herpesvirus 4"
/db_xref="taxon:10376"
<1..>288
/gene="EBNA-1"
<1..>288
/gene="EBNA-1"
/codon_start=1

```

```

/product="EBNA-1 protein variant P-ala'"
/protein_id="AAF23375.1"
/db_xref="GI:6683826"
/translation="GPSTGRGQGGRRKKGWFGKHRGCGSGNPKFENIAEGLRAL
LARSHVERTTEGTWAGVVFYGSKSTSLYNLRGVALAIPOCRLLPSRLP"
BASE COUNT 74 a 65 c 87 g 62 t
ORIGIN

Query Match 100.0%; Score 20; DB 14; Length 288;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ctccctttacaacctaaaggc 20
|||||
Db 213 CTCCTTTTACAACCTAAGGC 232

RESULT 4
HHU21195
LOCUS
DEFINITION Human herpesvirus 4, isolate FG nuclear antigen EBNA-1 gene,
partial cds, C-terminal unique domain.
ACCESSION U21195
VERSION U21195.1 GI:710375
KEYWORDS
SOURCE Epstein-Barr virus.
ORGANISM Human herpesvirus 4
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE 1 (bases 1 to 609)
AUTHORS Wright, M.N., Stewart, J.P., Janjua, N.J., Pepper, S.D., Sample, C.,
Rooney, C.M. and Arrand, J.R.
Antigenic and sequence variation in the C-terminal unique domain of
the Epstein-Barr virus nuclear antigen EBNA-1
Virology (1995) In press
2 (bases 1 to 627)
Pepper, S.D.
Direct Submission
Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for
Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,
Manchester, Britain, M20 9BX
FEATURES
source
Location/Qualifiers
1..627
/organism="Human herpesvirus 4"
/isolate="EG"
/specific_host="Homo sapiens"
/db_xref="taxon:10376"
<1..>612
/notes="C-terminal unique domain"
/codon_start=1
/product="nuclear antigen EBNA-1"
/protein_id="AAA67277.1"
/db_xref="GI:710376"
/translation="ADDPGEGPSTGPRGGDGGRRKKGWFGKHRGCGSGNPKFENIA
EGLRALLARSHVERTTDEGTWAGVVFYGSKSTSLYNLRGCTALAIPOCKLTPLSRLP
FGNAPGPGQPLRESIVCYFMVFLQTHIFAEVLKDAIKDLVMTKPAPTCNIRVTVC
SFDDGVDLPPFPFPMVEGAAGDGGDGGDGGEGEGQE"
BASE COUNT 154 a 132 c 195 g 146 t
ORIGIN

Query Match 100.0%; Score 20; DB 14; Length 627;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ctccctttacaacctaaaggc 20
|||||
Db 231 CTCCTTTTACAACCTAAGGC 250

RESULT 10
HHU21204

```

```
LOCUS HH021204 627 bp DNA linear VRL 24-MAY-1995
DEFINITION Human herpesvirus 4, isolate SB nuclear antigen EBNA-1 gene,
partial cds, C-terminal unique domain.
ACCESSION U21204
VERSION U21204.1 GI:710393
KEYWORDS
SOURCE Epstein-Barr virus.
ORGANISM Human herpesvirus 4
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE 1 (bases 1 to 609)
AUTHORS Wright, M.N., Stewart, J.P., Janjua, N.J., Pepper, S.D., Sample, C.,
Rooney, C.M. and Arrand, J.R.
TITLE Antigenic and sequence variation in the C-terminal unique domain of
the Epstein-Barr virus nuclear antigen EBNA-1
JOURNAL Virology (1995) In press
REFERENCE 2 (bases 1 to 627)
AUTHORS Pepper, S.D.
TITL Direct Submission
JOUR:AL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for
Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,
Manchester, Britain, M20 9BX
Location/Qualifiers
FEATURES
source
1..627
/organism="Human herpesvirus 4"
/isolate="SB"
/species="Homo sapiens"
/db_xref="taxon:10376"
<1..612
/note="C-terminal unique domain"
/codon_start=1
/product="nuclear antigen EBNA-1"
/protein_id="AA57286.1"
/translation="ADDPGEGSTGPRGGGRRKGGWFGKRGSGNPKPFENIA
DGRLLARSHVETTDGTVAGVYGGSKTSLNLRGTALAIPOCRLLPLSRLP
FGNAPGPGPQIPGVFMVFLQTHFAEVLKDAIKDLVMTKAPTNCIRVTVC
SFDDGVLDLPWFPMVFGAAAGDGDGDEGDEGEQ"
BASE COUNT 153 a 132 c 194 g 148 t
ORIGIN

Query Match 100.0%; Score 20; DB 14; Length 627;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctttacaacctaaggc 20
|||||
Db 231 CTCCTTTTACACCTAAGGC 250

RESULT 11
AR116267 AR116267 1748 bp DNA linear PAT 16-MAY-2001
LOCUS AR116267
DEFINITION Sequence 1 from patent US 6133025.
ACCESSION AR116267
VERSION AR116267.1 GI:14096589
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1748)
AUTHORS Seed, B.
TITL Compact Epstein-Barr virus replicons
JOURNAL Patent: US 6133025-A 1 17-OCT-2000;
FEATURES
source
1..1748
/organism="unknown"
BASE COUNT 453 a 382 c 546 g 367 t
ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 1748;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctttacaacctaaggc 20
|||||
Db 1928 CTCCTTTTACACCTAAGGC 1947

RESULT 14

LOCUS AX107940 1926 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 3 from Patent WO0125484.
ACCESSION AX107940
VERSION AX107940.1 GI:13923322
KEYWORDS
SOURCE Epstein-Barr virus.
ORGANISM Human herpesvirus 4
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE 1 (bases 1 to 1926)
AUTHORS Robertson, E.S. and Cotter, M.A.
TITLE Methods to inhibit or enhance the binding of viral dna to genomic
host dna
JOURNAL Patent: WO 0125484-A 3 12-APR-2001;
THE REGENTS OF THE UNIVERSITY OF MICHIGAN (US)
Location/Qualifiers
FEATURES
source
1..1926
/organism="Human herpesvirus 4"
/db_xref="taxon:10376"
BASE COUNT 487 a 352 c 872 g 215 t
ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 1926;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctttacaacctaaggc 20
|||||
Db 1545 CTCCTTTTACACCTAAGGC 1564

RESULT 13
AR108994 AR108994 2580 bp DNA linear PAT 14-FEB-2001
LOCUS AR108994
DEFINITION Sequence 2 from patent US 6114111.
ACCESSION AR108994
VERSION AR108994.1 GI:12825270
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2580)
AUTHORS Luo, Y., Huang, B. and Payan, D.
TITL Mammalian protein interaction cloning system
JOURNAL Patent: US 6114111-A 2 05-SEP-2000;
FEATURES
source
1..2580
/organism="unknown"
BASE COUNT 632 a 512 c 1054 g 382 t
ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 2580;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctccctttacaacctaaggc 20
|||||
Db 1928 CTCCTTTTACACCTAAGGC 1947

RESULT 14
```

AR083151/c  
LOCUS AR083151 5452 bp DNA linear PAT 01-SEP-2000  
DEFINITION Sequence 1 from patent US 5976807.  
ACCESSION AR083151  
VERSION AR083151.1 GI:10009941  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE  
AUTHORS 1 (bases 1 to 5452)  
TITLE Horlick, R.A., Dame B.B. and Robbins, A.K.  
JOURNAL Eukaryotic cells stably expressing genes from multiple transfected  
FEATURES  
source Patent: US 5976807-A 1 02-NOV-1999;  
Location/Qualifiers  
1..5452  
/organism="unknown"  
BASE COUNT 1108 a 1736 c 1246 g 1362 t  
ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 5452;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccctttacaacctaaaggc 20  
|||||  
Db 877 CTCCTTTACAACTAAGGC 858

RESULT 15  
U02454  
LOCUS U02454 5452 bp DNA circular SYN 29-MAR-1996  
DEFINITION Cloning vector pCMVBEA, complete sequence.  
ACCESSION U02454  
VERSION U02454.1 GI:413820  
KEYWORDS  
SOURCE Cloning vector pCMVBEA.  
ORGANISM Cloning vector pCMVBEA.  
REFERENCE 1 (bases 1 to 5452)  
AUTHORS Kitts, P.A.  
JOURNAL Unpublished  
TITLE ClONTECH Vectors On Disc version 1.3  
AUTHORS  
REFERENCE 2 (bases 1 to 5452)  
JOURNAL Swirski, R.A., Van Den Berg, D., Murphy, A.J., Lambert, C.M.,  
Friedberg, E.C. and Schimke, R.T.  
TITLE Improvements in the Epstein-Barr-based shuttle vector system for  
direct cloning in human tissue culture cells  
JOURNAL Methods: A Companion to Methods in Enzymology 4, 133-142 (1992)  
AUTHORS  
REFERENCE 3 (bases 1 to 5452)  
JOURNAL Kitts, P.A.  
TITLE Direct Submission  
JOURNAL Submitted (07-OCT-1993) Paul A. Kitts, CLONTECH Laboratories, Inc.,  
1020 East Meadow Circle, Palo Alto, CA 94303, USA  
COMMENT This vector can be obtained from CLONTECH Laboratories, Inc., 1020  
East Meadow Circle, Palo Alto, CA 94303, USA. To place an order  
call (415) 424-8222 or (800) 662-2566, extension 1. International  
customers, please contact your local distributor. For technical  
information, call (415) 424-8222 or (800) 662-2566, extension 3.  
This sequence has been compiled from information in the sequence  
databases, published literature and other sources, together with  
partial sequences obtained by CLONTECH; this vector has not been  
completely sequenced. If you suspect there is an error in this  
sequence, please contact CLONTECH's Technical Service Department at  
(415) 424-8222 or (800) 662-2566, extension 3 or E-mail  
TECH@CLONTECH.COM.  
FEATURES  
source Location/Qualifiers  
1..5452  
/organism="Cloning vector pCMVBEA"  
/db\_xref="taxon:31797"  
BASE COUNT 1362 a 1246 c 1736 g 1108 t  
ORIGIN

Query Match 100.0%; Score 20; DB 12; Length 5452;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccctttacaacctaaaggc 20  
|||||  
Db 2342 CTCCTTTACAACTAAGGC 2361

Search completed: May 18, 2002, 14:58:02  
Job time: 8179 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:26:24 ; Search time 250.21 Seconds  
(without alignments)  
24.543 Million cell updates/sec

Title: US-09-623-329-3  
Perfect score: 25  
Sequence: 1 agagacaaggtccttaatcgcatcc 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 ; Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	21	84.0	1748	3	US-09-100-730-1
C 2	21	84.0	2580	3	US-09-050-863-2
C 3	21	84.0	2580	4	US-09-359-081-2
C 4	21	84.0	5452	2	US-09-130-114-1
C 5	21	84.0	8575	5	PCT-US92-08258-6
C 6	21	84.0	9600	4	US-08-910-647-1
C 7	21	84.0	10596	1	US-07-884-811-15
C 8	21	84.0	10596	1	US-07-885-971-15
C 9	21	84.0	10596	1	US-08-087-783A-15
C 10	21	84.0	10596	1	US-08-194-088B-15
C 11	21	84.0	10596	2	US-08-194-087-15
C 12	21	84.0	10596	5	PCT-US93-04648-15
C 13	17.6	70.4	4615	2	US-08-674-351-3
C 14	16.8	67.2	146	2	US-08-743-200-7
C 15	16.8	67.2	2365	1	US-08-363-208-1
C 16	16.8	67.2	2365	4	US-09-137-478-1
C 17	16	64.0	1215	5	PCT-US96-05320A-707
C 18	16	64.0	1242	1	US-08-252-966B-13
C 19	16	64.0	1708	3	US-08-961-083-153
C 20	16	64.0	1738	4	US-08-858-207A-35
C 21	16	64.0	1774	3	US-08-961-083-73
C 22	16	64.0	1946	4	US-09-029-755C-4
C 23	16	64.0	2877	4	US-09-029-755C-3
C 24	15.8	63.2	1239	4	US-09-064-693A-20
C 25	15.8	63.2	4530	4	US-09-064-693A-26
C 26	15.6	62.4	339	1	US-08-322-742-6
C 27	15.6	62.4	492	1	US-08-009-261-3

C 28	15.6	62.4	2584	1	US-08-121-714-1	Sequence 1, Appli
C 29	15.6	62.4	2584	1	US-08-322-742-13	Sequence 13, Appl
C 30	15.6	62.4	2584	1	US-08-477-108A-1	Sequence 1, Appli
C 31	15.6	62.4	2584	2	US-08-477-112-1	Sequence 1, Appli
C 32	15.6	62.4	2584	5	PCT-US93-08322-1	Sequence 1, Appli
C 33	15.4	61.6	1125	3	US-09-306-881-1	Sequence 5, Appli
C 34	15.4	61.6	1700	3	US-08-860-519-5	Sequence 6, Appli
C 35	15.4	61.6	1700	3	US-08-860-519-6	Sequence 1, Appli
C 36	15.4	61.6	2350	4	US-09-187-478-1	Sequence 2, Appli
C 37	15.4	61.6	2488	1	US-08-279-270A-2	Sequence 6, Appli
C 38	15.4	61.6	3627	1	US-08-104-072B-6	Sequence 7, Appli
C 39	15.4	61.6	3627	1	US-08-351-413-7	Sequence 7, Appli
C 40	15.4	61.6	3627	2	US-09-025-583-7	Sequence 1, Appli
C 41	15.4	61.6	4527	3	US-09-045-632-1	Sequence 1, Appli
C 42	15.4	61.6	4716	2	US-08-290-731C-1	Sequence 10, Appl
C 43	15.2	60.8	300	3	US-08-965-903B-10	Sequence 30, Appl
C 44	15.2	60.8	1210	4	US-08-811-481-30	Sequence 9, Appli
C 45	15.2	60.8	1413	2	US-08-239-276-9	

## ALIGNMENTS

RESULT 1  
US-09-100-730-1/c  
; Sequence 1, Application US/09100730  
; Patent No. 6133025  
; GENERAL INFORMATION:  
; APPLICANT: Seed, Brian  
; TITLE OF INVENTION: COMPACT EPSTEIN-BARR VIRUS REPLICONS  
; FILE REFERENCE: 00786/357001  
; CURRENT APPLICATION NUMBER: US/09/100.730  
; CURRENT FILING DATE: 1998-06-19  
; EARLIER APPLICATION NUMBER: 60/050.206  
; EARLIER FILING DATE: 1997-06-19  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 1748  
; TYPE: DNA  
; ORGANISM: Epstein Barr Virus  
US-09-100-730-1

Query Match 84.0%; Score 21; DB 3; Length 1748;  
Best Local Similarity 100.0%; Pred. No. 0.064;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 acaaggtccttaatcgcatcc 25  
Db 1543 ACAAGGTCTTAATCGCATCC 1523

RESULT 2  
US-09-050-863-2/c  
; Sequence 2, Application US/09050863  
; Patent No. 6114111  
; GENERAL INFORMATION:  
; APPLICANT: Lao, Ying  
; APPLICANT: Hiang, Betty  
; APPLICANT: Payan, Don  
; TITLE OF INVENTION: Mammalian Protein Interaction Cloning  
; TITLE OF INVENTION: System  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Fiehr, Hohbach, Test, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM T.I.E: Floppy disk

; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/050,863  
; FILING DATE: 30-MAR-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silva, Robin M.  
; REGISTRATION NUMBER: 38,304  
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 949-8711  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2580 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA  
; US-09-050-863-2

Query Match 84.0%; Score 21; DB 3; Length 2580;  
Best Local Similarity 100.0%; Pred. No. 0.07;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaaggtccttaatcgatcc 25  
|||||  
Db 2130 ACAAGGTCCTTAATCGATCC 2110

RESULT 3  
US-09-359-081-2/c  
; Sequence 2, Application US/09359081  
; Patent No. 6316223  
; GENERAL INFORMATION:  
; APPLICANT: Lao, Ying  
; Hiang, Betty  
; Pavan, Don  
; TITLE OF INVENTION: Mammalian Protein Interaction Cloning  
; System  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fiehr, Hobbach, Test, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/359,081  
; FILING DATE: 22-Jul-1998  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/050,863  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silva, Robin M.  
; REGISTRATION NUMBER: 38,304  
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 949-8711  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2580 base pairs

; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-359-081-2

Query Match 84.0%; Score 21; DB 4; Length 2580;  
Best Local Similarity 100.0%; Pred. No. 0.07;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaaggtccttaatcgatcc 25  
|||||  
Db 2130 ACAAGGTCCTTAATCGATCC 2110

RESULT 4  
US-09-130-114-1  
; Sequence 1, Application US/09130114  
; Patent No. 5976807  
; GENERAL INFORMATION:  
; APPLICANT: Horlick, Robert A.  
; APPLICANT: Damaj, Bassam B.  
; APPLICANT: Robbins, Alan K.  
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes  
; FROM Multiple Transfected Episomes  
; FILE REFERENCE: 0867/1D903US1  
; CURRENT APPLICATION NUMBER: US/09/130,114  
; CURRENT FILING DATE: 1998-08-06  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 5452  
; TYPE: DNA  
; ORGANISM: VEBNA  
US-09-130-114-1

Query Match 84.0%; Score 21; DB 2; Length 5452;  
Best Local Similarity 100.0%; Pred. No. 0.085;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaaggtccttaatcgatcc 25  
|||||  
Db 674 acaaggtccttaatcgatcc 694

RESULT 5  
PCT-US92 08258-6/G  
; Sequence 6, Application PCT/US9208258  
; GENERAL INFORMATION:  
; APPLICANT: GENENTECH, INC. and STATE OF OREGON BY AND  
; TITLE OF INVENTION: PRODUCTION OF GPA NEUTROTROPIC FACTOR  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080-4990  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/08258  
; FILING DATE: 19920929  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:

;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Johnston, Sean A.  
;; REGISTRATION NUMBER: P35,910  
;; REFERENCE/DOCKET NUMBER: 731  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415/225,3562  
;; TELEFAX: 415/952-9881  
;; TELEX: 910/371-7168  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 8575 bases  
;; TYPE: NUCLEIC ACID  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
PCT-US92-08258-6

Query Match 84.0%; Score 21; DB 5; Length 8575;  
Best Local Similarity 100.0%; Pred. No. 0.095;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 acaaggtccttaatcgatcc 25  
|||||

Db 302 ACAAGGTCCTTAATCGATCC 282

RESULT 6  
US-08-910-647-1/c  
; Sequence 1, Application US/08910647  
; Patent No. 6251433  
; GENERAL INFORMATION:  
; APPLICANT: Zuckermann et al.  
; TITLE OF INVENTION: Compositions and Methods for  
; TITLE OF INVENTION: Polynucleotide Delivery  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 94608-2916  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/910,647  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fujita, Sharon M.  
; REGISTRATION NUMBER: 38,459  
; REFERENCE/DOCKET NUMBER: 1218,002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 923-2706  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9600 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-910-647-1

Query Match 84.0%; Score 21; DB 4; Length 9600;  
Best Local Similarity 100.0%; Pred. No. 0.097;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 acaaggtccttaatcgatcc 25  
|||||

Db 2177 ACAAGGTCCTTAATCGATCC 2157

RESULT 7  
US-07-884-811-15/c  
; Sequence 15, Application US/07884811  
; Patent No. 5316921  
; GENERAL INFORMATION:  
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.  
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: palin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/884,811  
; FILING DATE: 19920518  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Ginger R.  
; REGISTRATION NUMBER: 33,055  
; REFERENCE/DOCKET NUMBER: 755.1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-3216  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10596 bases  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-07-884-811-15

Query Match 84.0%; Score 21; DB 1; Length 10596;  
Best Local Similarity 100.0%; Pred. No. 0.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 acaaggtccttaatcgatcc 25  
|||||

Db 3711 ACAAGGTCCTTAATCGATCC 3691

RESULT 8  
US-07-885-971-15/c  
; Sequence 15, Application US/07885971  
; Patent No. 5328837  
; GENERAL INFORMATION:  
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.  
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/885,971  
FILING DATE: 19920518  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Ginger R.  
REGISTRATION NUMBER: 33,055  
REFERENCE/DOCKET NUMBER: 779  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-3216  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10596 bases  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-07-885-971-15

Query Match 84.0%; Score 21; DB 1; Length 10596;  
Best Local Similarity 100.0%; Pred. No. 0.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaaggtccttaatcgatcc 25  
|||||  
DB 3711 ACAAGGTCCTTAATCGCATCC 3691

RESULT 9  
US-08-087-783A-15/c  
Sequence 15, Application US/08087783A  
Patent No. 5547856  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.  
TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/087,783A  
FILING DATE: 13-Jul-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/884811  
FILING DATE: 18-MAY-92  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/885971  
FILING DATE: 18-MAY-92  
ATTORNEY/AGENT INFORMATION:  
NAME: Marschang, Diane L.  
REGISTRATION NUMBER: 35,600  
REFERENCE/DOCKET NUMBER: P0755779P1  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/225-5416  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10596 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-087-783A-15

Query Match 84.0%; Score 21; DB 1; Length 10596;  
Best Local Similarity 100.0%; Pred. No. 0.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaaggtccttaatcgatcc 25  
|||||  
DB 3711 ACAAGGTCCTTAATCGCATCC 3691

RESULT 10  
US-08-194-088B-15/C  
Sequence 15, Application US/08194088B  
Patent No. 5580963  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.  
TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/194,088B  
FILING DATE: 09-FEB-1994  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/884811  
FILING DATE: 18-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Gallegos, R. Thomas  
REGISTRATION NUMBER: 42,692  
REFERENCE/DOCKET NUMBER: 75,501  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-2614  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10596 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-194-088B-15

Query Match 84.0%; Score 21; DB 1; Length 10596;  
Best Local Similarity 100.0%; Pred. No. 0.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaaggtccttaatcgatcc 25  
|||||  
DB 3711 ACAAGGTCCTTAATCGCATCC 3691

```
RESULT 11
JS-08-194-087-15/c
; Sequence 15, Application US/08194087
; Patent No. 5879910
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,087
; FILING DATE: 18-MAY-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-194-087-15

Query Match 84.0%; Score 21; DB 2; Length 10596;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaaggtccttaatcgatcc 25
|||||
DB 3711 ACAAGGTCCTTAATCGCATCC 3691

RESULT 12
PCT-US93-04648-15/c
; Sequence 15, Application PC/TUS9304648
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc., Godowski, Paul J., Lokker, Natalie A., Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/04648
; FILING DATE: 19930517
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/884811
; FILING DATE: 18-MAY-92
; PRIOR APPLICATION DATA: 07/885971
; APPLICATION NUMBER:
; FILING DATE: 18-MAY-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 755,779P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; PCT-US93-04648-15

Query Match 84.0%; Score 21; DB 5; Length 10596;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaaggtccttaatcgatcc 25
|||||
DB 3711 ACAAGGTCCTTAATCGCATCC 3691

RESULT 13
US-08-674-351-3/c
; Sequence 3, Application US/08674351
; Patent No. 5831013
; GENERAL INFORMATION:
; APPLICANT: Bruenn, Jeremy A.
; APPLICANT: Yao, Wensheng
; TITLE OF INVENTION: CAPSID POLYPEPTIDES AND USE TO INHIBIT
; TITLE OF INVENTION: VIRAL PACKAGING
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hararave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/674,351
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Timian, Susan J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 19226/740
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1636
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4615 base pairs
; TYPE: nucleic acid
```



; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
US-08-674-351-3

Query Match 70.4%; Score 17.6; DB 2; Length 4615;  
Best Local Similarity 83.3%; Pred. No. 5;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 agagacaaggtccttaatgcaccc 24  
||| ||||| | ||||| |||||  
DB 833 AGACACAGGATCGTAATCGCATC 810

## RESULT 14

US-08-743-200-7/c

; Sequence 7, Application US/08743200  
; Patent No. 5861260  
; GENERAL INFORMATION:  
; APPLICANT: Doxsey, Stephen J.  
; TITLE OF INVENTION: DIAGNOSTIC METHODS FOR SCREENING  
; PATIENTS FOR SCLERODERMA  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08743.200  
; FILING DATE: 05-NOV-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fasse, J. Peter  
; REGISTRATION NUMBER: 32,983  
; REFERENCE/DOCKET NUMBER: 07917/025001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-542-5070  
; TELEFAX: 617-542-8906  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 146 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: Coding Sequence  
; LOCATION: 1...144

US-08-743-200-7

Query Match 67.2%; Score 16.8; DB 2; Length 146;  
Best Local Similarity 90.0%; Pred. No. 5.7;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 caaggtccttaatgcaccc 25  
||||| ||||| ||||| |||||  
DB 118 CAAGGTCTTAATCTCTCC 99

## RESULT 15

US-08-363-208-1/c

; Sequence 1, Application US/08363208  
; Patent No. 5767366  
; GENERAL INFORMATION:  
; APPLICANT: Sathasivan, Kanagasabapathi  
; APPLICANT: Murai, No. 5767366imoto  
; TITLE OF INVENTION: A Mutant Acetolactate Synthase Gene From  
; Arabidopsis Thaliana For Conferring Imidazolinone  
; TITLE OF INVENTION: Resistance To Crop Plants  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Llewellyn A. Proctor, Sr.  
; STREET: 11481 Sheraton Drive  
; CITY: Baton Rouge,  
; STATE: LA  
; COUNTRY: USA  
; ZIP: 70815  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/363.208  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/657,429  
; FILING DATE: 19-FEB-1991  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Proctor Sr., Llewellyn A.  
; REGISTRATION NUMBER: 20,152  
; REFERENCE/DOCKET NUMBER: 013911-001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (504)275-8689  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2365 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-363-208-1

Query Match 67.2%; Score 16.8; DB 1; Length 2365;  
Best Local Similarity 90.0%; Pred. No. 11;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 caaggtccttaatgcaccc 25  
||||| ||||| ||||| |||||  
DB 1722 CAAGGACCTTAATCGCATAC 1703

Search completed: May 18, 2002, 17:26:27  
Job time: 12829 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 14:58:02 ; Search time 2878.96 Seconds  
(without alignments)  
181.720 Million cell updates/sec

Title: US-09-623-329-3

Perfect score: 25

Sequence: 1 agagacaaggtccttaacgcaccc 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank :

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vt.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vt.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
-----						

1	25	100.0	25	6	AX018444	AX018444 Sequence
2	21.4	85.6	627	14	HHU21194	U21194 Human herpe
3	21.4	85.6	627	14	HHU21196	U21196 Human herpe
4	21.4	85.6	627	14	HHU21201	U21201 Human herpe
5	21	84.0	36	6	BD001005	BD001005 Method an
6	21	84.0	36	6	BD001434	BD001434 Method an
7	21	84.0	627	14	HHU21193	U21193 Human herpe
8	21	84.0	627	14	HHU21195	U21195 Human herpe
9	21	84.0	627	14	HHU21197	U21197 Human herpe
10	21	84.0	627	14	HHU21198	U21198 Human herpe
11	21	84.0	627	14	HHU21199	U21199 Human herpe
12	21	84.0	627	14	HHU21200	U21200 Human herpe
13	21	84.0	627	14	HHU21202	U21202 Human herpe
14	21	84.0	627	14	HHU21203	U21203 Human herpe
15	21	84.0	627	14	HHU21204	U21204 Human herpe
16	21	84.0	627	14	HHU21205	U21205 Human herpe
17	21	84.0	636	14	HHU21205	U21205 Human herpe
18	21	84.0	1748	6	AR116267	AR116267 Sequence
19	21	84.0	1926	6	AX107940	AX107940 Sequence
20	21	84.0	2580	6	AR108994	AR108994 Sequence
21	21	84.0	5452	6	AR083151	AR083151 Sequence
22	21	84.0	5452	12	U02454	U02454 Cloning vec
23	21	84.0	9600	6	A92665	A92665 Sequence 1
24	21	84.0	9600	6	AR158345	AR158345 Sequence
25	21	84.0	10596	6	I25041	I25041 Sequence 15
26	21	84.0	10596	6	I30503	I30503 Sequence 15
27	21	84.0	10737	12	XX002428	U02428 Cloning vec
28	21	84.0	10850	12	U02455	U02455 Cloning vec
29	21	84.0	11265	6	AX031159	AX031159 Sequence
30	21	84.0	17281	14	EBV	V01555 Epstein-Bar
31	20.2	80.8	184113	14	HS4B958RAJ	M80517 Epstein-Bar
32	19.8	79.2	39316	2	AC022878	AC022878 Homo sapi
33	19.8	79.2	39316	2	AC014425	AC014425 Drosophil
34	19.4	77.6	252370	3	AE003529	AE003529 Drosophil
35	19.4	77.6	110000	2	AC015817_0	AC015817 Homo sapi
36	19.4	77.6	110000	2	AC015817_2	Continuation (3 of
37	19.4	77.6	169162	2	AC012479	AC012479 Homo sapi
38	19.4	77.6	194032	2	AC019150	AC019150 Homo sapi
39	19.4	77.6	195799	2	AC087742	AC087742 Homo sapi
40	19.2	76.8	195392	2	AC087292	AC087292 Homo sapi
41	18.8	75.2	83798	9	HS1013A22	AL132772 Human DNA
42	18.8	75.2	189563	2	AC106128	AC106128 Rattus no
43	18.8	75.2	193742	2	AL669932	AL669932 Mus muscu
44	18.6	74.4	246502	2	AL669828	AL669828 Mus muscu
45	18.6	74.4	408	3	KCSD8B8	X75829 H.contortus
			2258	4	SSC293583	AJ293583 Sus scrofa

## ALIGNMENTS

RESULT	1	AX018444	Sequence 3 from Patent WO9945155.	25 bp	DNA	linear	PAT 07-SEP-2000
LOCUS	AX018444	Sequence 3 from Patent WO9945155.					
DEFINITION	AX018444	Sequence 3 from Patent WO9945155.					
ACCESSION	AX018444	Sequence 3 from Patent WO9945155.					
VERSION	AX018444.1	GI:10042595					
KEYWORDS							
SOURCE							
ORGANISM							

Epstein-Barr virus.  
Human herpesvirus 4  
Viruses; dsDNA viruses, no RNA stage: Herpesviridae;  
Gammaherpesvirinae; Lymphocryptovirus.  
1 (bases 1 to 25)

Middelkoop, J.M., Van Den Brule, A.J. and Vervoot, M.B.  
Oligonucleotides for the amplification and detection of Epstein  
Barr virus (ebv) nucleic acid  
Patent: WO 9945155-A 3 10-SEP-1999;  
MIDDELDORP JAAP MICHEL (NL); AKZO NOBEL NV (NL); DEN BRULE  
ADRIANUS JOHANNES CH (NL); VERVOORT MARCEL BARTOLINA HEND (NL)

FEATURES  
Location/Qualifiers  
1..25  
/organism="Human herpesvirus 4"  
/db\_xref="taxon:10376"

BASE COUNT 8 a 7 c 5 g 5 t



```

CDS
/specfic_host="Homo sapiens"
/db_xref="taxon:10376"
<1_612
/Note="C-terminal unique domain"
/codon_start=1
/product="nuclear antigen EBNA-1"
/protein_id="AA67283.1"
/db_xref="GI:710388"
/translation="ADDPGEGPSTGPGGGRKKGWFGKRRGGSGNPKPENIA
EGLRLLLARHCHVETDGNVAGVYVGGSKTSLYLRGGIGLAIPOCRILTPLSLRP
FOMAPGPGQGLRESYCYFVFLQTHFAEGLKDAIKDLVLPKPAPTCLNIRVTVC
SFDDGVLPFPFPMVGGAAEGDDGDDGDEGDEGEQE"
BASE COUNT 151 a 129 c 196 g 151 t
ORIGIN

Query Match 85.6%; Score 21.4; DB 14; Length 627;
Best Local Similarity 95.7%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 agacaaggtccttaacgcattcc 25
|||||
DB 436 AACAGGTCCTTAATCGCATCC 414

RESULT 5
BD001005/c
LOCUS
DEFINITION Method and reagent for inhibiting viral replication.
ACCESSION BD001005
VERSION BD001005.1 GI:18625564
KEYWORDS JP 2000342285-A/165.
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 36)
AUTHORS Draper, K.G., Dadykiz, L.W., Macswigen, J.A., Maysejak, D.G.,
Holesek, J.J., and Mamone, A.J.
TITLE Method and reagent for inhibiting viral replication
JOURNAL RIBOZYME PHARMACEUTICALS INC
COMMENT OS Artificial Sequence
PN JP 2000342285-A/165
PD 12-DEC-2000
PF 01-MAY-2000 JP 2000132616
PR 11-MAY-1992 US 07/882689,14-MAY-1992 US 07/882712 PR
14-MAY-1992 US 07/882823,14-MAY-1992 US 07/882824 PR
14-MAY-1992 US 07/882886,14-MAY-1992 US 07/882888 PR
14-MAY-1992 US 07/882889,14-MAY-1992 US 07/882921 PR
14-MAY-1992 US 07/882922,14-MAY-1992 US 07/883823 PR
14-MAY-1992 US 07/883849,14-MAY-1992 US 07/884073 PR
14-MAY-1992 US 07/884074,14-MAY-1992 US 07/884333 PR
14-MAY-1992 US 07/884422,14-MAY-1992 US 07/884431 PR
14-MAY-1992 US 07/884436,14-MAY-1992 US 07/884521 PR
31-JUL-1992 US 07/923738,26-AUG-1992 US 07/935854 PR
26-AUG-1992 US 07/936086,18-SEP-1992 US 07/948359 PR
15-OCT-1992 US 07/963322,07-DEC-1992 US 07/987129 PR
07-DEC-1992 US 07/987130,07-DEC-1992 US 07/987133 PR
KENNETH G DRAPER,LEC W DADYKIZ,JAMES A MACSWIGEN, PI DENNIS G
MAYSEJAK,
PI JAMES J HOLESEK,ANTHONY J MAMONE
PC C12N15/09,C12N5/10,C12N7/00,C12N9/22,A61K38/43,A61K39/125,A61K39/13,
A61K39/135,
PC A61K39/145,A61K39/21,A61K39/23,A61K39/245,A61K39/29,A61K48/00,
A61P1/16,
PC A61P31/14,A61P31/16,A61P31/18,A61P31/22,A61P35/02,C12Q1/68,PC
(C12N15/09,C12R1:93),C12N15/00,C12N5/ ,A61K37/48,(C12N15/00,PC
C12R1:93)
CC Key Location/Qualifiers
FH source 1..36
FT Location/Qualifiers
FEATURES
source
1..36
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 12 a 6 c 10 g 8 t
ORIGIN

Query Match 84.0%; Score 21; DB 6; Length 36;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaaggtccttaacgcattcc 25
|||||
DB 24 ACAAGGTCCTTAATCGCATCC 4

RESULT 6
BD001434/c
LOCUS
DEFINITION Method and reagent for inhibiting viral replication.
ACCESSION BD001434
VERSION BD001434.1 GI:18625993
KEYWORDS JP 2000342286-A/165.
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (bases 1 to 36)
AUTHORS Draper, K.G., Dadykiz, L.W., Macswigen, J.A., Maysejak, D.G.,
Holesek, J.J., and Mamone, A.J.
TITLE Method and reagent for inhibiting viral replication
JOURNAL RIBOZYME PHARMACEUTICALS INC
COMMENT OS Artificial Sequence
PN JP 2000342286-A/165
PD 12-DEC-2000
PF 01-MAY-2000 JP 2000132651
PR 11-MAY-1992 US 07/882689,14-MAY-1992 US 07/882712 PR
14-MAY-1992 US 07/882823,14-MAY-1992 US 07/882824 PR
14-MAY-1992 US 07/882886,14-MAY-1992 US 07/882888 PR
14-MAY-1992 US 07/882889,14-MAY-1992 US 07/882921 PR
14-MAY-1992 US 07/882922,14-MAY-1992 US 07/883823 PR
14-MAY-1992 US 07/883849,14-MAY-1992 US 07/884073 PR
14-MAY-1992 US 07/884074,14-MAY-1992 US 07/884333 PR
14-MAY-1992 US 07/884422,14-MAY-1992 US 07/884431 PR
14-MAY-1992 US 07/884436,14-MAY-1992 US 07/884521 PR
31-JUL-1992 US 07/923738,26-AUG-1992 US 07/935854 PR
26-AUG-1992 US 07/936086,18-SEP-1992 US 07/948359 PR
15-OCT-1992 US 07/963322,07-DEC-1992 US 07/987129 PR
07-DEC-1992 US 07/987130,07-DEC-1992 US 07/987133 PR
KENNETH G DRAPER,LEC W DADYKIZ,JAMES A MACSWIGEN, PI DENNIS G
MAYSEJAK,
PI JAMES J HOLESEK,ANTHONY J MAMONE
PC C12N15/09,C12N5/10,C12N7/00,A61K38/43,A61K39/125,A61K39/13,
A61K39/135,
PC A61K39/145,A61K39/21,A61K39/23,A61K39/245,A61K39/29,A61K48/00,
A61P1/16,
PC A61P31/14,A61P31/16,A61P31/18,A61P31/22,A61P35/02,C12Q1/68,PC
(C12N15/09,C12R1:93),C12N15/00,C12N5/ ,A61K37/48,(C12N15/00,PC
C12R1:93)
CC Key Location/Qualifiers
FH source 1..36
FT Location/Qualifiers
FEATURES
source
1..36
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 12 a 6 c 10 g 8 t
ORIGIN

Query Match 84.0%; Score 21; DB 6; Length 36;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



LOCUS  
DEFINITION Human herpesvirus 4, isolate LA nuclear antigen EBNA-1 gene,  
partial cds, C-terminal unique domain.  
ACCESSION U21198.1 GI:710381  
VERSION 1  
KEYWORDS  
SOURCE Epstein-Barr virus.  
ORGANISM Human herpesvirus 4  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Gammaherpesvirinae; Lymphocryptovirus.  
REFERENCE 1 (bases 1 to 609)  
AUTHORS Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,  
Rooney,C.M. and Arrand,J.R.  
TITLE Antigenic and sequence variation in the C-terminal unique domain of  
the Epstein-Barr virus nuclear antigen EBNA-1  
JOURNAL Virology (1995) In press  
REFERENCE 2 (bases 1 to 627)  
AUTHORS Pepper,S.D.  
TITLE Direct Submission  
JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for  
Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,  
Manchester, Britain, M20 9BX

Query Match 84.0%; Score 21; DB 14; Length 627;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 acaaggtccttaacgcaccc 25  
|||||  
Db 434 ACAAGGTCCTTAATCGCATCC 414

RESULT 10  
LOCUS HHU21198/c  
DEFINITION Human herpesvirus 4, isolate LA nuclear antigen EBNA-1 gene,  
partial cds, C-terminal unique domain.  
ACCESSION U21198.1 GI:710381  
VERSION 1  
KEYWORDS  
SOURCE Epstein-Barr virus.  
ORGANISM Human herpesvirus 4  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Gammaherpesvirinae; Lymphocryptovirus.  
REFERENCE 1 (bases 1 to 609)  
AUTHORS Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,  
Rooney,C.M. and Arrand,J.R.  
TITLE Antigenic and sequence variation in the C-terminal unique domain of  
the Epstein-Barr virus nuclear antigen EBNA-1  
JOURNAL Virology (1995) In press  
REFERENCE 2 (bases 1 to 627)  
AUTHORS Pepper,S.D.  
TITLE Direct Submission  
JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for  
Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,  
Manchester, Britain, M20 9BX

FEATURES  
source  
1. .627  
/organism="Human herpesvirus 4"  
/isolate="LA"  
/specific\_host="Homo sapiens"  
/db\_xref="taxon:10376"  
<1..612  
/note="C-terminal unique domain"  
/codon\_start=1  
/product="nuclear antigen EBNA-1"  
/protein\_id="AAA67280.1"  
/db\_xref="GI:710381"

REFERENCE 1 (bases 1 to 609)  
AUTHORS Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,  
Rooney,C.M. and Arrand,J.R.  
TITLE Antigenic and sequence variation in the C-terminal unique domain of  
the Epstein-Barr virus nuclear antigen EBNA-1  
JOURNAL Virology (1995) In press  
REFERENCE 2 (bases 1 to 627)  
AUTHORS Pepper,S.D.  
TITLE Direct Submission  
JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for  
Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,  
Manchester, Britain, M20 9BX

BASE COUNT 155 a 132 c 191 g 149 t  
ORIGIN

Query Match 84.0%; Score 21; DB 14; Length 627;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 acaaggtccttaacgcaccc 25  
|||||  
Db 434 ACAAGGTCCTTAATCGCATCC 414

RESULT 11  
LOCUS HHU21199/c  
DEFINITION Human herpesvirus 4, isolate NL nuclear antigen EBNA-1 gene,  
partial cds, C-terminal unique domain.  
ACCESSION U21200.1 GI:710385  
VERSION 1  
KEYWORDS  
SOURCE Epstein-Barr virus.  
ORGANISM Human herpesvirus 4  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Gammaherpesvirinae; Lymphocryptovirus.  
REFERENCE 1 (bases 1 to 609)  
AUTHORS Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,  
Rooney,C.M. and Arrand,J.R.  
TITLE Antigenic and sequence variation in the C-terminal unique domain of  
the Epstein-Barr virus nuclear antigen EBNA-1  
JOURNAL Virology (1995) In press  
REFERENCE 2 (bases 1 to 627)  
AUTHORS Pepper,S.D.  
TITLE Direct Submission  
JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for

Query Match 84.0%; Score 21; DB 14; Length 627;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 acaaggtccttaacgcaccc 25  
|||||  
Db 434 ACAAGGTCCTTAATCGCATCC 414

RESULT 11  
LOCUS HHU21199/c

LOCUS HHU21199  
DEFINITION Human herpesvirus 4, isolate MT nuclear antigen EBNA-1 gene,  
partial cds, C-terminal unique domain.  
ACCESSION U21199.1 GI:710383  
VERSION 1  
KEYWORDS  
SOURCE Epstein-Barr virus.  
ORGANISM Human herpesvirus 4  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Gammaherpesvirinae; Lymphocryptovirus.  
REFERENCE 1 (bases 1 to 609)  
AUTHORS Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,  
Rooney,C.M. and Arrand,J.R.  
TITLE Antigenic and sequence variation in the C-terminal unique domain of  
the Epstein-Barr virus nuclear antigen EBNA-1  
JOURNAL Virology (1995) In press  
REFERENCE 2 (bases 1 to 627)  
AUTHORS Pepper,S.D.  
TITLE Direct Submission  
JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for  
Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,  
Manchester, Britain, M20 9BX

FEATURES  
source  
1. .627  
/organism="Human herpesvirus 4"  
/isolate="MT"  
/specific\_host="Homo sapiens"  
/db\_xref="taxon:10376"  
<1..612  
/note="C-terminal unique domain"  
/codon\_start=1  
/product="nuclear antigen EBNA-1"  
/protein\_id="AAA67281.1"  
/db\_xref="GI:710384"

REFERENCE 1 (bases 1 to 609)  
AUTHORS Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,  
Rooney,C.M. and Arrand,J.R.  
TITLE Antigenic and sequence variation in the C-terminal unique domain of  
the Epstein-Barr virus nuclear antigen EBNA-1  
JOURNAL Virology (1995) In press  
REFERENCE 2 (bases 1 to 627)  
AUTHORS Pepper,S.D.  
TITLE Direct Submission  
JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for  
Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,  
Manchester, Britain, M20 9BX

BASE COUNT 153 a 131 c 193 g 150 t  
ORIGIN

Query Match 84.0%; Score 21; DB 14; Length 627;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 acaaggtccttaacgcaccc 25  
|||||  
Db 434 ACAAGGTCCTTAATCGCATCC 414

RESULT 12  
LOCUS HHU21200/c  
DEFINITION Human herpesvirus 4, isolate NL nuclear antigen EBNA-1 gene,  
partial cds, C-terminal unique domain.  
ACCESSION U21200.1 GI:710385  
VERSION 1  
KEYWORDS  
SOURCE Epstein-Barr virus.  
ORGANISM Human herpesvirus 4  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Gammaherpesvirinae; Lymphocryptovirus.  
REFERENCE 1 (bases 1 to 609)  
AUTHORS Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,  
Rooney,C.M. and Arrand,J.R.  
TITLE Antigenic and sequence variation in the C-terminal unique domain of  
the Epstein-Barr virus nuclear antigen EBNA-1  
JOURNAL Virology (1995) In press  
REFERENCE 2 (bases 1 to 627)  
AUTHORS Pepper,S.D.  
TITLE Direct Submission  
JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for

BASE COUNT 153 a 131 c 193 g 150 t  
ORIGIN

Query Match 84.0%; Score 21; DB 14; Length 627;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 acaaggtccttaacgcaccc 25  
|||||  
Db 434 ACAAGGTCCTTAATCGCATCC 414

RESULT 12  
LOCUS HHU21200/c  
DEFINITION Human herpesvirus 4, isolate NL nuclear antigen EBNA-1 gene,  
partial cds, C-terminal unique domain.  
ACCESSION U21200.1 GI:710385  
VERSION 1  
KEYWORDS  
SOURCE Epstein-Barr virus.  
ORGANISM Human herpesvirus 4  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Gammaherpesvirinae; Lymphocryptovirus.  
REFERENCE 1 (bases 1 to 609)  
AUTHORS Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,  
Rooney,C.M. and Arrand,J.R.  
TITLE Antigenic and sequence variation in the C-terminal unique domain of  
the Epstein-Barr virus nuclear antigen EBNA-1  
JOURNAL Virology (1995) In press  
REFERENCE 2 (bases 1 to 627)  
AUTHORS Pepper,S.D.  
TITLE Direct Submission  
JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for

BASE COUNT 155 a 132 c 191 g 149 t  
ORIGIN

Query Match 84.0%; Score 21; DB 14; Length 627;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 acaaggtccttaacgcaccc 25  
|||||  
Db 434 ACAAGGTCCTTAATCGCATCC 414

RESULT 11  
LOCUS HHU21199/c

Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,  
Manchester, Britain, M20 9BX

# FEATURES

Source  
1..627  
/organism="Human herpesvirus 4"  
/isolate="NL"  
/specific\_host="Homo sapiens"  
/db\_xref="taxon:10376"  
<1..612  
/note="C-terminal unique domain"  
/codon\_start=1  
/product="nuclear antigen EBNA-1"  
/protein\_id="AAA67282.1"  
/db\_xref="GI:710386"

## CDS

EGRLTLARSHVETTDGTVAGVYGGSKTSLNLRGIALAIPQCRLTPLSRLP  
FGMAPGPGOPGLRESIVCYFIVLQTHIFAEGLKDAIKDLVLPKPAFTCNIRVTC  
SFDDGVDLPPWFPMPVEGAAAGDDGDDGDEGDEGEQGE"  
BASE COUNT 154 a 133 c 192 g 148 t  
ORIGIN

Query Match 84.0%; Score 21; DB 14; Length 627;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaaggtccttaatcgatcc 25

|||||

Db 434 ACAAGGTCCTTAATCGATCC 414

## RESULT 13

HHU21202/c  
LOCUS  
DEFINITION Human herpesvirus 4, isolate PA nuclear antigen EBNA-1 gene,  
partial cds, C-terminal unique domain.  
U21202  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Epstein-Barr virus.

Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

Gammaherpesvirinae; Lymphocryptovirus.

1 (bases 1 to 609)

Wrightam,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,

Rooney,C.M. and Arrand,J.R.

Antigenic and sequence variation in the C-terminal unique domain of

the Epstein-Barr virus nuclear antigen EBNA-1

virology (1995) In press

2 (bases 1 to 627)

Pepper,S.D.

Direct Submission

Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for

Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,

Manchester, Britain, M20 9BX

## FEATURES

Source  
1..627  
/organism="Human herpesvirus 4"  
/isolate="PA"  
/specific\_host="Homo sapiens"  
/db\_xref="taxon:10376"  
<1..612  
/note="C-terminal unique domain"  
/codon\_start=1  
/product="nuclear antigen EBNA-1"  
/protein\_id="AAA67284.1"  
/db\_xref="GI:710390"

## CDS

EGRLTLARSHVETTDGTVAGVYGGSKTSLNLRGIALAIPQCRLTPLSRLP  
FGMAPGPGOPGLRESIVCYFIVLQTHIFAEGLKDAIKDLVLPKPAFTCNIRVTC  
SFDDGVDLPPWFPMPVEGAAAGDDGDDGDEGDEGEQGE"  
BASE COUNT 155 a 130 c 193 g 149 t  
ORIGIN

Query Match 84.0%; Score 21; DB 14; Length 627;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaaggtccttaatcgatcc 25

|||||

Db 434 ACAAGGTCCTTAATCGATCC 414

## RESULT 14

HHU21203/c  
LOCUS  
DEFINITION Human herpesvirus 4, isolate PM nuclear antigen EBNA-1 gene,  
partial cds, C-terminal unique domain.  
U21203  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Epstein-Barr virus.

Human herpesvirus 4

Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

Gammaherpesvirinae; Lymphocryptovirus.

1 (bases 1 to 609)

Wrightam,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,

Rooney,C.M. and Arrand,J.R.

Antigenic and sequence variation in the C-terminal unique domain of

the Epstein-Barr virus nuclear antigen EBNA-1

virology (1995) In press

2 (bases 1 to 627)

Pepper,S.D.

Direct Submission

Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for

Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,

Manchester, Britain, M20 9BX

## FEATURES

Source  
1..627  
/organism="Human herpesvirus 4"  
/isolate="PM"  
/specific\_host="Homo sapiens"  
/db\_xref="taxon:10376"  
<1..612  
/note="C-terminal unique domain"  
/codon\_start=1  
/product="nuclear antigen EBNA-1"  
/protein\_id="AAA67285.1"  
/db\_xref="GI:710392"

## CDS

EGRLTLARSHVETTDGTVAGVYGGSKTSLNLRGIALAIPQCRLTPLSRLP  
FGMAPGPGOPGLRESIVCYFIVLQTHIFAEGLKDAIKDLVLPKPAFTCNIRVTC  
SFDDGVDLPPWFPMPVEGAAAGDDGDDGDEGDEGEQGE"  
BASE COUNT 155 a 132 c 192 g 148 t  
ORIGIN

Query Match 84.0%; Score 21; DB 14; Length 627;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 acaaggtccttaatcgatcc 25

|||||

Db 434 ACAAGGTCCTTAATCGATCC 414

## RESULT 15

HHU21204/c  
LOCUS  
DEFINITION Human herpesvirus 4, isolate SB nuclear antigen EBNA-1 gene,  
partial cds, C-terminal unique domain.  
U21204  
VERSION  
KEYWORDS  
SOURCE

Epstein-Barr virus.

ORGANISM Human herpesvirus 4  
 Viruses; dsDNA viruses, no RNA stage: Herpesviridae;  
 Gammaherpesvirinae; Lymphocryptovirus.  
 REFERENCE 1 (bases 1 to 609)  
 AUTHORS Wright, M.N., Stewart, J.P., Janjua, N.J., Pepper, S.D., Sample, C.,  
 Rooney, C.M. and Arrand, J.R.  
 TITLE Antigenic and sequence variation in the C-terminal unique domain of  
 the Epstein-Barr virus nuclear antigen EBNA-1  
 JOURNAL Virology (1995) In press  
 REFERENCE 2 (bases 1 to 627)  
 AUTHORS Pepper, S.D.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for  
 Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,  
 Manchester, Britain, M20 9BX  
 FEATURES  
 source  
 1..627  
 /organism="Human herpesvirus 4"  
 /isolate="SB"  
 /specific\_host="Homo sapiens"  
 /db\_xref="taxon:10376"  
 <1..612  
 /note="C-terminal unique domain"  
 /codon\_start=1  
 /product="nuclear antigen EBNA-1"  
 /protein\_id="AAA67286.1"  
 /db\_xref="GI:710394"  
 /translation="ADDPGEPTGPRGGKGGWFGKURGGGSGNPKFENIA  
 DGLRALLARSHVERTDEGTWAGVFTGGSKTLYNIIRGTALAIPOCRLLTPLSLP  
 FGMAPGPGPQPLRESICVFMVFLQTHIFAELKDAIKDLVMTKPAPTCNIRVTVC  
 SFDDGVDLPPWFPPMVEGAAAGDDGDDGDEGDEGDEG"  
 BASE COUNT 153 a 132 c 194 g 148 t  
 ORIGIN

Query Match 84.0%; Score 21; DB 14; Length 627;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 5 acaaggtccttaatcgcatcc 25  
 ||||||||||||||||  
 Db 434 ACAAGTCTTAAATCGCATCC 414

Search completed: May 18, 2002, 14:58:03  
 Job time: 8180 sec





; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/050,863  
; FILING DATE: 30-MAR-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silva, Robin M.  
; REGISTRATION NUMBER: 38,304  
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 949-8711  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2580 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA  
; US-09-050-863-2

Query Match 100.0%; Score 20; DB 3; Length 2580;  
Best Local Similarity 100.0%; Pred. No. 0.1;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccctttacaacctaaagc 20  
|||||  
DB 1928 CTCCTTTTACACCTAAGGC 1947

## RESULT 3

; Sequence 2, Application US/09359081  
; Patent No. 6316223  
; GENERAL INFORMATION:  
; APPLICANT: Lao, Ying

; Hiang, Betty  
; Pavan, Don

; TITLE OF INVENTION: Mammalian Protein Interaction Cloning  
; SYSTEM

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94111-4187

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA: US/09/359,081

; APPLICATION NUMBER: US/09/359,081

; FILING DATE: 22-Jul-1999

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/050,863

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Silva, Robin M.

; REGISTRATION NUMBER: 38,304

; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 781-1989

; TELEFAX: (415) 949-8711

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2580 base pairs

; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-359-081-2

Query Match 100.0%; Score 20; DB 4; Length 2580;  
Best Local Similarity 100.0%; Pred. No. 0.1;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccctttacaacctaaagc 20  
|||||  
DB 1928 CTCCTTTTACACCTAAGGC 1947

## RESULT 4

; Sequence 1, Application US/09130114  
; Patent No. 5976807  
; GENERAL INFORMATION:  
; APPLICANT: Horlick, Robert A.

; Damaj, Bassam B.

; APPLICANT: Robbins, Alan K.

; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes

; TITLE OF INVENTION: From Multiple Transfected Episomes

; FILE REFERENCE: 0867/ID903US1

; CURRENT APPLICATION NUMBER: US/09/130,114

; CURRENT FILING DATE: 1998-08-06

; NUMBER OF SEQ ID NOS: 36

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 5452

; TYPE: DNA

; ORGANISM: VEBNA

; US-09-130-114-1

Query Match 100.0%; Score 20; DB 2; Length 5452;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccctttacaacctaaagc 20  
|||||  
DB 877 CTCCTTTTACACCTAAGGC 858

## RESULT 5

; Sequence 6, Application PC/TUS9208258

; GENERAL INFORMATION:

; APPLICANT: GENENTECH, INC. and STATE OF OREGON BY AND

; TITLE OF INVENTION: PRODUCTION OF GPA NEUROTROPIC FACTOR

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 460 Point San Bruno Blvd

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 9408014990

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: patin (Genentech)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US92/08258

; FILING DATE: 19920929

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Johnston, Sean A.  
;; REGISTRATION NUMBER: P35,910  
;; REFERENCE/DOCKET NUMBER: 731  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415/2253562  
;; TELEFAX: 415/952-9881  
;; TELEX: 910/371-7168  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 8575 bases  
;; TYPE: NUCLEIC ACID  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
PCT-US92-08258-6

Query Match 100.0%; Score 20; DB 5; Length 8575;  
Best Local Similarity 100.0%; Pred. No. 0.12; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0;

Qy 1 ctccctttacaaacctaaagc 20  
Db 99 CTCCTTTTACAACTTAAGGC 118

RESULT 6  
US-08-910-647-1  
; Sequence 1, Application US/08910647  
; Patent No. 6251433  
; GENERAL INFORMATION:  
; APPLICANT: Zuckermann et al.  
; TITLE OF INVENTION: Compositions and Methods for  
; TITLE OF INVENTION: Polynucleotide Delivery  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 94608-2916  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/910,647  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fujita, Sharon M.  
; REGISTRATION NUMBER: 38,459  
; REFERENCE/DOCKET NUMBER: 1218,002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 923-2706  
; TELEFAX: (510) 655-3542  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9600 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-910-647-1

Query Match 100.0%; Score 20; DB 4; Length 9600;  
Best Local Similarity 100.0%; Pred. No. 0.12; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0;

Qy 1 ctccctttacaaacctaaagc 20  
Db 1974 CTCCTTTTACAACTTAAGGC 1993

RESULT 7  
US-07-884-811-15  
; Sequence 15, Application US/07884811  
; Patent No. 5316921  
; GENERAL INFORMATION:  
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.  
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/884,811  
; FILING DATE: 19920518  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Ginger R.  
; REGISTRATION NUMBER: 33,055  
; REFERENCE/DOCKET NUMBER: 755,1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-3216  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10596 bases  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-07-884-811-15

Query Match 100.0%; Score 20; DB 1; Length 10596;  
Best Local Similarity 100.0%; Pred. No. 0.12; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 0;

Qy 1 ctccctttacaaacctaaagc 20  
Db 3508 CTCCTTTTACAACTTAAGGC 3527

RESULT 8  
US-07-885-971-15  
; Sequence 15, Application US/07885971  
; Patent No. 5328837  
; GENERAL INFORMATION:  
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.  
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/885.971  
FILING DATE: 19920518  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dregger, Ginger R.  
REGISTRATION NUMBER: 33,055  
REFERENCE/DOCKET NUMBER: 779  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-3216  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10596 bases  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-07-885-971-15

Query Match 100.0%; Score 20; DB 1; Length 10596;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccctttacaacctaaagc 20  
DB 3508 CTCCTTTACACCTAAGC 3527

RESULT 9  
US-08-087-783A-15  
Sequence 15, Application US/08087783A  
Patent No. 5547856  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.  
TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/087.783A  
FILING DATE: 13-Jul-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/884811  
FILING DATE: 18-MAY-92  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/885971  
FILING DATE: 18-MAY-92  
ATTORNEY/AGENT INFORMATION:  
NAME: Marschang, Diane L.  
REGISTRATION NUMBER: 35,600  
REFERENCE/DOCKET NUMBER: P0755779P1  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/225-5416  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10596 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-087-783A-15

Query Match 100.0%; Score 20; DB 1; Length 10596;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccctttacaacctaaagc 20  
DB 3508 CTCCTTTACACCTAAGC 3527

RESULT 10  
US-08-194-088B-15  
Sequence 15, Application US/08194088B  
Patent No. 5580963  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.  
TITLE OF INVENTION: SINGLE-CHAIN EPITOPIC GROWTH FACTOR VARIANTS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/194.088B  
FILING DATE: 09-FEB-1994  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/884811  
FILING DATE: 18-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Gallegos, R. Thomas  
REGISTRATION NUMBER: 32,692  
REFERENCE/DOCKET NUMBER: 75501  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-2614  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10596 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-194-088B-15

Query Match 100.0%; Score 20; DB 1; Length 10596;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccctttacaacctaaagc 20  
DB 3508 CTCCTTTACACCTAAGC 3527

```
RESULT 11
US-08-194-087-15
; Sequence 15, Application US/08194087
; Patent No. 5879910
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194.087
; FILING DATE: 18-MAY-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-194-087-15

Query Match 100.0%; Score 20; DB 2; Length 10596;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccctttacaacctaagc 20
Db 3508 CTCCCTTACAACCTAAGC 3527

RESULT 12
PCT-US93-04648-15
; Sequence 15, Application PC/TUS9304648
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc., Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998.416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/04648
; FILING DATE: 19930517
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/884811
; FILING DATE: 18-MAY-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/885971
; FILING DATE: 18-MAY-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 755,779P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
PCT-US93-04648-15

Query Match 100.0%; Score 20; DB 5; Length 10596;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctccctttacaacctaagc 20
Db 3508 CTCCCTTACAACCTAAGC 3527

RESULT 13
US-08-998-416-594/c
; Sequence 594, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippssen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Redischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPHII
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998.416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
```

REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 594:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 571 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1407UP  
US-08-998-416-594

Query Match 72.0%; Score 14.4; DB 4; Length 571;  
Best Local Similarity 93.8%; Pred. No. 71;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0:  
QY 5 ctttacacctaaggc 20  
||||| |||||  
Db 523 CTTTACAGCCTAAGC 508

RESULT 14  
US-08-222-124-10  
Sequence 10, Application US/08222124  
Patent No. 5620881  
GENERAL INFORMATION:  
APPLICANT: Wieder, Kenneth J., Strom,  
APPLICANT: Terry B.  
TITLE OF INVENTION: Gene Encoding Mutant L3T4  
TITLE OF INVENTION: Protein Which Facilitates  
TITLE OF INVENTION: HIV Infection and  
TITLE OF INVENTION: Transgenic Mouse Expressing  
TITLE OF INVENTION: Such Protein  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)  
SOFTWARE: Wordperfect (Version 5.0)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/222,124  
FILING DATE: 04-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/046,025  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Paul T. Clark  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 01948/022003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1408  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-222-124-10

Query Match 71.0%; Score 14.2; DB 1; Length 1408;  
Best Local Similarity 84.2%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0:  
QY 1 ctccctttacaacctaagg 19  
||||| |  
Db 549 CTCCATGTCCAACCTAAGG 567  
RESULT 15  
US-08-842-657A-10  
Sequence 10, Application US/08842657A  
Patent No. 5936138  
GENERAL INFORMATION:  
APPLICANT: Wieder, Kenneth J., Strom,  
APPLICANT: Terry B.  
TITLE OF INVENTION: Gene Encoding Mutant L3T4  
TITLE OF INVENTION: Protein Which Facilitates  
TITLE OF INVENTION: HIV Infection and  
TITLE OF INVENTION: Transgenic Mouse Expressing  
TITLE OF INVENTION: Such Protein  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)  
SOFTWARE: Wordperfect (Version 5.0)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/842,657A  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/222,124  
FILING DATE: 04-APR-1995  
APPLICATION NUMBER: 08/046,025  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Paul T. Clark  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 01948/022003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1408  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-842-657A-10

Query Match 71.0%; Score 14.2; DB 2; Length 1408;  
Best Local Similarity 84.2%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0:  
QY 1 ctccctttacaacctaagg 19  
||||| |  
Db 549 CTCCATGTCCAACCTAAGG 567

Search completed: May 18, 2002, 17:26:24  
Job time: 12826 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

COM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:26:29 ; Search time 250.21 Seconds  
(without alignments)

Title: US-09-623-329-5  
Perfect score: 30  
Sequence: 1 cgtctccctttggaatgccctggacc 30

Scoring table: IDENTITY\_NUC  
Gapop 10.0 : Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*
```

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	30	100.0	1748	3	US-09-100-730-1		Sequence 1, Appli
2	30	100.0	2580	3	US-09-050-863-2		Sequence 2, Appli
3	30	100.0	2580	4	US-09-359-081-2		Sequence 2, Appli
c 4	30	100.0	5452	2	US-09-130-114-1		Sequence 6, Appli
5	30	100.0	8575	5	PCT-US92-08258-6		Sequence 1, Appli
6	30	100.0	9600	4	US-08-910-647-1		Sequence 1, Appli
7	30	100.0	10596	1	US-07-884-811-15		Sequence 15, Appli
8	30	100.0	10596	1	US-07-885-971-15		Sequence 15, Appli
9	30	100.0	10596	1	US-08-087-783A-15		Sequence 15, Appli
10	30	100.0	10596	1	US-08-194-088B-15		Sequence 15, Appli
11	30	100.0	10596	1	US-08-194-087-15		Sequence 15, Appli
12	30	100.0	10596	5	PCT-US93-04648-15		Sequence 15, Appli
13	18.8	62.7	1164	1	US-08-445-640-5		Sequence 5, Appli
14	18.8	62.7	1164	3	US-08-170-558-5		Sequence 5, Appli
15	18.8	62.7	1164	3	US-08-447-314-5		Sequence 5, Appli
16	18.8	62.7	1164	3	US-08-445-461-5		Sequence 5, Appli
17	18.8	62.7	2742	1	US-08-286-305A-2		Sequence 2, Appli
18	18.8	62.7	2742	2	US-08-441-104A-2		Sequence 2, Appli
19	18.8	62.7	2742	2	US-08-440-816A-2		Sequence 2, Appli
20	18.8	62.7	2742	4	US-09-417-381A-2		Sequence 2, Appli
21	18.8	62.7	3611	1	US-08-445-640-1		Sequence 1, Appli
22	18.8	62.7	3611	3	US-08-170-558-1		Sequence 1, Appli
23	18.8	62.7	3611	3	US-08-447-314-1		Sequence 1, Appli
24	18.8	62.7	3611	3	US-08-445-461-1		Sequence 1, Appli
c 25	17.8	59.3	927	1	US-08-479-173-7		Sequence 7, Appli
c 26	17.8	59.3	927	1	US-08-479-173-8		Sequence 8, Appli
c 27	17.8	59.3	927	1	US-08-340-732-1		Sequence 1, Appli

C	28	17.8	59.3	927	4	US-08-880-131A-7	Sequence 7, Appl	
		17.8	59.3	927	4	US-08-880-131A-8	Sequence 8, Appl	
	C	29	17.8	59.3	927	4	US-08-864-641B-1	Sequence 1, Appl
	C	30	17.8	59.3	927	4	US-09-668-822-7	Sequence 7, Appl
	C	31	17.8	59.3	927	4	US-09-668-822-8	Sequence 8, Appl
	C	32	17.8	59.3	927	4	PCT-US95-14952-1	Sequence 1, Appl
	C	33	17.8	59.3	927	5	US-08-249-687C-1	Sequence 1, Appl
		34	17.8	59.3	4975	2	US-08-666-392A-3	Sequence 3, Appl
		35	17.8	59.3	4989	2	US-08-625-819-1	Sequence 1, Appl
		36	17.8	59.3	4989	3	US-08-755-558-4	Sequence 4, Appl
		37	17.8	59.3	4989	3	US-08-746-559A-1	Sequence 1, Appl
		38	17.8	59.3	4989	4	US-08-880-131A-9	Sequence 9, Appl
		39	17.8	59.3	4989	4	US-09-199-926-3	Sequence 3, Appl
		40	17.8	59.3	4989	4	US-08-864-641B-17	Sequence 17, Appl
		41	17.8	59.3	4989	4	US-09-668-822-9	Sequence 9, Appl
		42	17.8	59.3	4989	4	US-08-746-559A-3	Sequence 3, Appl
		43	17.8	59.3	4989	3	US-08-696-139-1	Sequence 1, Appl
	C	44	17.4	58.0	1215	1	US-09-136-652-1	Sequence 1, Appl
		45	17.4	58.0	1215	3		

## ALIGNMENTS

```

RESULT      1
US-09-100-730-1
: Sequence 1, Application US/09100730
: Patent No. 6133025
: GENERAL INFORMATION:
: APPLICANT: Seed, Brian
: TITLE OF INVENTION: COMPACT EPSTEIN-BARR VIRUS REPLICONS
: FILE REFERENCE: 00786/357001
: CURRENT APPLICATION NUMBER: US/09/100,730
: CURRENT FILING DATE: 1998-06-19
: EARLIER APPLICATION NUMBER: 60/050,206
: EARLIER FILING DATE: 1997-06-19
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 1748
: TYPE: DNA
: ORGANISM: Epstein Barr Virus
US-09-100-730-1

```

```
Query Match      100.0%; Score 30; DB 3; Length 1748;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 30: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

1 cgtctcccccttgggaatggccccctggacc 30  
+++++  
1407 cgtctcccccttgggaatggccccctggacc 1436

```

RESULT          2
US-09-050-863-2
: Sequence 2, Application US/09050863
: Patent No. 611411
: GENERAL INFORMATION:
: APPLICANT: Lao, Ying
: APPLICANT: Hiang, Betty
: APPLICANT: Payan, Don
: TITLE OF INVENTION: Mammalian Protein Interaction Cloning
: TITLE OF INVENTION: System
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/050,863  
FILING DATE: 30-MAR-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 949-8711  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2580 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
US-09-050-863-2

Query Match 100.0%; Score 30; DB 3; Length 2580;  
Best Local Similarity 100.0%; Pred. No. 0.00012;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgtctcccttgggaatgccccctggacc 30  
|||||  
DB 1995 CGTCTCCCTTTGGAATGGCCCTGGACCC 2024

## RESULT 3

US-09-359-081-2  
Sequence 2, Application US/09359081  
Patent No. 6316223  
GENERAL INFORMATION:  
APPLICANT: Lao, Ying  
Huang, Betty  
Payan, Don  
TITLE OF INVENTION: Mammalian Protein Interaction Cloning  
System  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-4187

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/359,081  
FILING DATE: 22-Jul-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/050,863  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 949-8711  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2580 base pairs

TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-359-081-2  
Query Match 100.0%; Score 30; DB 4; Length 2580;  
Best Local Similarity 100.0%; Pred. No. 0.00012;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgtctcccttgggaatgccccctggacc 30  
|||||  
DB 1995 CGTCTCCCTTTGGAATGGCCCTGGACCC 2024

## RESULT 4

US-09-130-114-1/c  
Sequence 1, Application US/09130114  
Patent No. 5976807  
GENERAL INFORMATION:  
APPLICANT: Horlick, Robert A.  
APPLICANT: Dama, Hassam B.  
APPLICANT: Robbins, Alan K.  
TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes  
FROM MULTIPLE TRANSFECTED EPISOMES  
FILE REFERENCE: 0867/ID903US1  
CURRENT APPLICATION NUMBER: US/09/130,111  
CURRENT FILING DATE: 1998-08-06  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 5452  
TYPE: DNA  
ORGANISM: VEBNA  
US-09-130-114-1

Query Match 100.0%; Score 30; DB 2; Length 5452;  
Best Local Similarity 100.0%; Pred. No. 0.00013;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgtctcccttgggaatgccccctggacc 30  
|||||  
DB 810 CGTCTCCCTTTGGAATGGCCCTGGACCC 781

## RESULT 5

PCT-US92-08258-6  
Sequence 6, Application PC/TUS9208258  
GENERAL INFORMATION:  
APPLICANT: GENENTECH, INC. and STATE OF OREGON BY AND  
TITLE OF INVENTION: PRODUCTION OF GPA NEUROTROPIC FACTOR  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080-4990  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/08258  
FILING DATE: 19920929  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:



;; FILING DATE: 100.0%; Score 30; DB 5; Length 8575;  
;; ATTORNEY/AGENT INFORMATION: 100.0%; Pred. No. 0.00014;  
;; NAME: Johnston, Sean A. Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
;; REGISTRATION NUMBER: P35,910  
;; REFERENCE/DOCKET NUMBER: 731  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415/2253562  
;; TELEFAX: 415/952-9881  
;; TELEX: 910/371-7168  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 8575 bases  
;; TYPE: NUCLEIC ACID  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; PCT-US92-08258-6

Query Match 100.0%; Score 30; DB 5; Length 8575;  
Best Local Similarity 100.0%; Pred. No. 0.00014;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgtctcccttgggaatgcccctggacc 30  
DB 166 CGTCTCCCTTTGGGAATGCCCTGGACCC 195

RESULT 6  
US-08-910-647-1  
;; Sequence 1, Application US/08910647  
;; Patent No. 6251433  
;; GENERAL INFORMATION:  
;; APPLICANT: zuckermann et al.  
;; TITLE OF INVENTION: Compositions and Methods for  
;; TITLE OF INVENTION: Polynucleotide Delivery  
;; NUMBER OF SEQUENCES: 4  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Chiron Corporation  
;; STREET: 4560 Horton Street  
;; CITY: Emeryville  
;; STATE: California  
;; COUNTRY: U.S.A.  
;; ZIP: 94608-2916  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent in Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; FILING DATE:  
;; CLASSIFICATION: 514  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Fujita, Sharon M.  
;; REGISTRATION NUMBER: 38,459  
;; REFERENCE/DOCKET NUMBER: 1218.002  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (510) 923-2706  
;; TELEFAX: (510) 655-3542  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 9600 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
US-08-910-647-1

Query Match 100.0%; Score 30; DB 4; Length 9600;  
Best Local Similarity 100.0%; Pred. No. 0.00014;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgtctcccttgggaatgcccctggacc 30  
DB 2041 CGTCTCCCTTTGGGAATGCCCTGGACCC 2070

RESULT 7  
US-07-884-811-15  
;; Sequence 15, Application US/07884811  
;; Patent No. 5316921  
;; GENERAL INFORMATION:  
;; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.  
;; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS  
;; NUMBER OF SEQUENCES: 21  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Genentech, Inc.  
;; STREET: 460 Point San Bruno Blvd  
;; CITY: South San Francisco  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 94080  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: patin (Genentech)  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/884,811  
;; FILING DATE: 19920518  
;; CLASSIFICATION: 530  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER:  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Dreger, Ginger R.  
;; REGISTRATION NUMBER: 33,055  
;; REFERENCE/DOCKET NUMBER: 755.1  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 415/225-3216  
;; TELEFAX: 415/952-9881  
;; TELEX: 910/371-7168  
;; INFORMATION FOR SEQ ID NO: 15:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 10596 bases  
;; TYPE: NUCLEIC ACID  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
US-07-884-811-15

Query Match 100.0%; Score 30; DB 1; Length 10596;  
Best Local Similarity 100.0%; Pred. No. 0.00015;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgtctcccttgggaatgcccctggacc 30  
DB 3575 CGTCTCCCTTTGGGAATGCCCTGGACCC 3604

RESULT 8  
US-07-885-971-15  
;; Sequence 15, Application US/07885971  
;; Patent No. 5328837  
;; GENERAL INFORMATION:  
;; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.  
;; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS  
;; NUMBER OF SEQUENCES: 21  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Genentech, Inc.  
;; STREET: 460 Point San Bruno Blvd  
;; CITY: South San Francisco  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 94080

Query Match 100.0%; Score 30; DB 4; Length 9600;  
Best Local Similarity 100.0%; Pred. No. 0.00014;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,971
; FILING DATE: 19920518
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-07-885-971-15

Query Match 100.0%; Score 30; DB 1; Length 10596;
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgtctcccttgggaatggccctggacc 30
Db 3575 CGTCTCCCTTTGGAATGGCCCTGGACCC 3604

RESULT 10
US-08-194-088B-15
; Sequence 15, Application US/08194088B
; Patent No. 5580963
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,088B
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/884811
; FILING DATE: 18-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallegos, R. Thomas
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: 75501
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-2614
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-194-088B-15

Query Match 100.0%; Score 30; DB 1; Length 10596;
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cgtctcccttgggaatggccctggacc 30
Db 3575 CGTCTCCCTTTGGAATGGCCCTGGACCC 3604

US-08-087-783A-15
; Sequence 15, Application US/08087783A
; Patent No. 5547856
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/087,783A
; FILING DATE: 13-Jul-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/884811
; FILING DATE: 18-MAY-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Marschang, Diane L.
; REGISTRATION NUMBER: 35,600
; REFERENCE/DOCKET NUMBER: P0755779P1
; TELECOMMUNICATION INFORMATION:
```

;/ CURRENT APPLICATION DATA:  
;/ APPLICATION NUMBER: PCT/US93/04648  
;/ FILING DATE: 19930517  
;/ CLASSIFICATION:  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: 07/884811  
;/ FILING DATE: 18-MAY-92  
;/ PRIOR APPLICATION DATA:  
;/ APPLICATION NUMBER: 07/885971  
;/ FILING DATE: 18-MAY-92  
;/ ATTORNEY/AGENT INFORMATION:  
;/ NAME: Dreger, Ginger R.  
;/ REGISTRATION NUMBER: 33,055  
;/ REFERENCE/DOCKET NUMBER: 755,779p1  
;/ TELECOMMUNICATION INFORMATION:  
;/ TELEPHONE: 415/225-3216  
;/ TELEFAX: 415/952-9881  
;/ TELEX: 910/371-7168  
;/ INFORMATION FOR SEQ ID NO: 15:  
;/ SEQUENCE CHARACTERISTICS:  
;/ LENGTH: 10596 bases  
;/ TYPE: NUCLEIC ACID  
;/ STRANDEDNESS: single  
;/ TOPOLOGY: linear  
;/ PCT-US93-04648-15

Query Match 100.0%; Score 30; DB 5; Length 10596;  
Best Local Similarity 100.0%; Pred. No. 0.00015;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgtctcccttgggaatgccccctggacc 30  
Db 3575 CGTCTCCCTTTGGGAATGCCCTGGACCC 3604

RESULT 13  
US-08-445-640-5  
; Sequence 5, Application US/08445640  
; Patent No. 5709858  
; GENERAL INFORMATION:  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Mark, Melanie R.  
; APPLICANT: Scadden, David T.  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Baron, Will F.  
; TITLE OF INVENTION: Protein Tyrosine Kinases  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/445,640  
; FILING DATE: 22-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/170558  
; FILING DATE: 20-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/157563  
; FILING DATE: 23-NOV-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hasak, Janet E.  
; REGISTRATION NUMBER: 28,616

RESULT 11  
US-08-194-087-15  
; Sequence 15, Application US/08194087  
; Patent No. 5879910  
; GENERAL INFORMATION:  
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.  
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/194,087  
; FILING DATE: 18-MAY-1992  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Ginger R.  
; REGISTRATION NUMBER: 33,055  
; REFERENCE/DOCKET NUMBER: 779  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-3216  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10596 bases  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-194-087-15

Query Match 100.0%; Score 30; DB 2; Length 10596;  
Best Local Similarity 100.0%; Pred. No. 0.00015;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 cgtctcccttgggaatgccccctggacc 30  
Db 3575 CGTCTCCCTTTGGGAATGCCCTGGACCC 3604

RESULT 12  
PCT-US93-04648-15  
; Sequence 15, Application PC/TUS9304648  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc., Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.  
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)

REFERENCE/DOCKET NUMBER: 854C2  
TELEPHONE: 415/225-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1164 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-445-640-5

Query Match 62.7%; Score 18.8; DB 1; Length 1164;  
Best Local Similarity 76.7%; Pred. No. 10;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 cgtctcccttgggaatggccctggacc 30  
|| ||||| ||||| ||||| ||  
Db 921 CGAGGCCCTTTGGAGGCCCTTGGGACC 950

RESULT 14  
US-08-170-558-5  
Sequence 5, Application US/08170558  
Patent No. 6001621  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J.  
APPLICANT: Mark, Melanie R.  
APPLICANT: Scadden, David T.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Baron, Will F.  
TITLE OF INVENTION: Protein Tyrosine Kinases  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/170,558  
FILING DATE: 20-DEC-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/157563  
FILING DATE: 23-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 854C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1164 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-170-558-5

Query Match 62.7%; Score 18.8; DB 3; Length 1164;  
Best Local Similarity 76.7%; Pred. No. 10;

Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
Qy 1 cgtctcccttgggaatggccctggacc 30  
|| ||||| ||||| ||||| ||  
Db 921 CGAGGCCCTTTGGAGGCCCTTGGGACC 950

RESULT 15  
US-08-447-314-5  
Sequence 5, Application US/08447314  
Patent No. 6087144  
GENERAL INFORMATION:  
APPLICANT: Scadden, David T.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Baron, Will F.  
TITLE OF INVENTION: Protein Tyrosine Kinases  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/447,314  
FILING DATE: 22-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/170558  
FILING DATE: 20-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/157563  
FILING DATE: 23-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 854C1D2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1164 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-447-314-5

Query Match 62.7%; Score 18.8; DB 3; Length 1164;  
Best Local Similarity 76.7%; Pred. No. 10;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 cgtctcccttgggaatggccctggacc 30  
|| ||||| ||||| ||||| ||  
Db 921 CGAGGCCCTTTGGAGGCCCTTGGGACC 950

Search completed: May 18, 2002, 17:26:32  
Job time: 12834 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: May 18, 2002, 13:56:51 ; Search time 1024.22 seconds  
(without alignments)  
33.526 Million cell updates/sec

Title: US-09-623-329-2  
Perfect score: 20  
Sequence: 1 ctcccttaccactaagc 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /SID55/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID55/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID55/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID55/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID55/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID55/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID55/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID55/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID55/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID55/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID55/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID55/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID55/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID55/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID55/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID55/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	AAZ11652	Oligo specific for
2	20	100.0	20	AAZ11652	PCR-generated eEBN
3	20	100.0	1092	AAZ11652	Epstein-Barr nucle
4	20	100.0	1212	AAZ11652	Epstein-Barr antigen
5	20	100.0	1748	AAZ11652	Epstein-Barr virus
6	20	100.0	1925	AAZ11652	Epstein Barr virus
7	20	100.0	1926	AAZ11652	Epstein Barr virus
8	20	100.0	1926	AAZ11652	EBV tethering prot
9	20	100.0	2580	AAZ11652	Nucleotide sequenc

c	10	20	100.0	5452	20	AAZ11652	Anti-sense strand
c	11	20	100.0	8575	14	AAZ11652	Sequence of pHEBO3
c	12	20	100.0	8705	10	AAZ11652	Vector pShuttle DN
c	13	20	100.0	9600	19	AAZ11652	Vector plasmid pCM
c	14	20	100.0	9725	21	AAZ11652	Expression vector
c	15	20	100.0	9732	21	AAZ11652	Expression vector
c	16	20	100.0	9738	21	AAZ11652	Expression vector
c	17	20	100.0	9873	21	AAZ11652	Expression vector
c	18	20	100.0	10054	21	AAZ11652	Expression vector
c	19	20	100.0	10380	20	AAZ11652	Nucleotide sequenc
c	20	20	100.0	10596	14	AAZ11652	Plasmid pCisEBON f
c	21	20	100.0	10596	17	AAZ11652	Plasmid pCisEBON f
c	22	20	100.0	10596	20	AAZ11652	Nucleotide sequenc
c	23	20	100.0	11265	19	AAZ11652	Plasmid pREP7::CRL
c	24	18.4	92.0	16080	21	AAZ11652	DNA clone pCEK C1.
c	25	17	85.0	70	21	AAZ11652	Human secreted pro
c	26	15.8	79.0	374	16	AAZ11652	Human gene signatu
c	27	15.8	79.0	401	22	AAZ11652	Human neuregulin 9
c	28	15.8	79.0	401	22	AAZ11652	Human neuregulin 9
c	29	15.8	79.0	2025	22	AAZ11652	Human cDNA sequenc
c	30	15.8	79.0	2150	21	AAZ11652	Human secreted pro
c	31	15.8	79.0	2241	22	AAZ11652	Human cDNA sequenc
c	32	15.8	79.0	2363	22	AAZ11652	Human cDNA sequenc
c	33	15.8	79.0	5801	22	AAZ11652	Human immune/haema
c	34	15.8	79.0	9692	22	AAZ11652	DNA encoding novel
c	35	15.8	79.0	9730	22	AAZ11652	DNA encoding novel
c	36	15.8	79.0	11348	22	AAZ11652	Human immune/haema
c	37	15.8	79.0	11350	22	AAZ11652	Human immune/haema
c	38	15.8	79.0	1503900	22	AAZ11652	Human neuregulin-1
c	39	15.8	79.0	1503900	22	AAZ11652	Human neuregulin-1
c	40	15.4	77.0	1916	18	AAZ11652	Brevibacterium fla
c	41	15.4	77.0	2199	22	AAZ11652	C glutamic codin
c	42	15.4	77.0	2435	23	AAZ11652	DNA encoding novel
c	43	15.4	77.0	2544	21	AAZ11652	Human nucleic acid
c	44	15.4	77.0	2841	24	AAZ11652	Human immune syste
c	45	15.4	77.0	4943	23	AAZ11652	Drosophila melanog

## ALIGNMENTS

RESULT	1
ID	AAZ11652 standard; DNA; 20 BP.
AC	AAZ11652;
DT	19-NOV-1999 (first entry)
DE	Oligo specific for Epstein Barr Nuclear Antigen 1 (EBNA-1) RNA.
XX	Epstein Barr Virus; EBV infection; viral; gene transcription; EBER-1;
KW	Epstein Barr early RNA; Epstein Barr nuclear antigen 1; EBNA-1; LMP-1;
KW	latent membrane protein; LMP-2; vIL10; BCRF-1; BARF1; BDLF2; NASBA;
KW	EBV-associated malignancy; primer; ss.
OS	Synthetic.
OS	Epstein-barr virus.
PN	W09945155-A2.
XX	10-SEP-1999.
XX	01-MAR-1999; 99WO-EP01392.
XX	04-MAR-1998; 98EP-0200655.
XX	14-DEC-1998; 98EP-0204231.
PA	(ALKU ) AKZO NOBEL NV.
XX	Vervooort MBHJ, Van Den Brule AJC, Middelorp JM;
XX	WPI; 1999-551051/46.

XX Identifying Epstein Barr Virus infection -  
 PT Claim 11; Page 20; 50pp; English.  
 PS  
 XX The invention provides methods for identifying an Epstein Barr Virus  
 CC (EBV) infection, that comprises determining viral gene transcription  
 CC patterns by amplification of specific RNA sequences. The binding sites  
 CC of the oligos suitable for amplification are located in the following  
 CC genes: Epstein Barr early RNA (EBER-1), Epstein Barr nuclear antigen 1  
 CC (EBNA-1), latent membrane protein 1 (LMP-1), LMP-2, and villo (BCRF-1),  
 CC BARF1 and BDLF2. The method comprises (a) amplifying a target sequence  
 CC within one or more RNA(s) transcribed from above gene sequences and the  
 CC (b) detecting the amplified products, determining the transcription  
 CC pattern and identifying the corresponding EBV-associated malignancy. The  
 CC RNA is amplified using a transcription based amplification technique  
 CC such as NASBA. The invention is used to diagnose malignant and  
 CC non-malignant EBV-associated diseases. Sequences AA211651-54 represent  
 CC oligos specific for EBNA-1 RNA.  
 XX  
 SQ Sequence 20 BP; 5 A; 8 C; 2 G; 5 T; 0 other;

Query Match 100.0%; Score 20; DB 20; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 0.39;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctcccttacaacctaaggc 20  
 |||||  
 Db 1 ctcccttacaacctaaggc 20

## RESULT 2

AAT15397  
 ID AAT15397 standard; DNA; 624 BP.  
 XX  
 AC AAT15397;  
 XX  
 DT 19-JUL-1996 (first entry)  
 XX  
 DE PCR-generated eBNAL protein gene.  
 XX  
 KW EBV; nuclear antigen protein 1; EBNA1; immunoassay; ds.  
 XX  
 OS Epstein-barr virus.  
 XX  
 FH Key Location/Qualifiers  
 FT primer\_bind complement (1..24)  
 FT /\*tag= a  
 FT /note= "primer for eBNAL"  
 FT 584..624  
 FT /\*tag= b  
 FT /note= "primer for eBNAL"  
 XX  
 PN WO9602563-A1.  
 XX  
 PD 01-FEB-1996.  
 XX  
 PF 13-JUL-1995; 95WO-US08700.  
 XX  
 PR 13-JUL-1994; 94US-0275614.  
 XX  
 PA (CORR ) CORNELL RES FOUND INC.  
 XX  
 PI O'Donnell ME;  
 XX  
 DR WPI; 1996-105853/11.  
 DR P-PSDB; AAR88588.  
 XX  
 PT Recovery of Epstein Barr Virus nuclear antigen protein 1 (EBNA1) -  
 PT useful in immunoassay system for detecting Epstein Barr Virus in  
 PT serum samples  
 XX

## Example 22; Page 51; 82pp; English.

PS  
 XX A DNA fragment (AAT15397) was generated by PCR amplification of  
 CC the Epstein-Barr virus (EBV) nuclear antigen protein 1 (EBNA1)  
 CC gene (see also AAT15396) in plasmid p291 and NdeI/BamHI digestion of  
 CC the PCR product. It was ligated into pET3c and the resulting  
 CC plasmid was used to transform Escherichia coli strain BL21  
 CC (DE3)pLysS cells. The recombinant EBNA1, designated eBNAL  
 CC (AAR88588), can be recovered from the nuclei of host cells for use  
 XX EBV immunoassays.  
 SQ Sequence 624 BP; 148 A; 132 C; 195 G; 149 T; 0 other;

Query Match 100.0%; Score 20; DB 17; Length 624;  
 Best Local Similarity 100.0%; Pred. No. 0.56;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ctcccttacaacctaaggc 20  
 |||||  
 Db 222 ctcccttacaacctaaggc 241

RESULT 3  
 AAX03478  
 ID AAX03478 standard; DNA; 1092 BP.  
 XX  
 AC AAX03478;  
 XX  
 DT 26-APR-1999 (first entry)  
 XX  
 DE Epstein-Barr nuclear antigen 1 (EBNA1) coding region.  
 XX  
 KW EBV; replicon; episomal replication; Orip; EBNA1;  
 KW Epstein-Barr nuclear antigen 1; gene therapy; vector; ss.  
 XX  
 OS Epstein-Barr virus.  
 XX  
 PN WO9857658-A1.  
 XX  
 PD 23-DEC-1998.  
 XX  
 PF 14-APR-1998; 98WO-US07641.  
 XX  
 PR 19-JUN-1997; 97US-0050206.  
 XX  
 PA (GEHO ) GEN HOSPITAL CORP.  
 XX  
 PI Seed B;  
 XX  
 DR WPI; 1999-080933/07.  
 XX  
 PT Epstein-Barr virus replicons which support episomal replication -  
 PT comprise an Orip sequence and an EBNA1 sequence operably linked to a  
 PT promoter  
 XX  
 PS Claim 4; Page -; 12pp; English.  
 XX  
 CC This is the Epstein-Barr nuclear antigen 1 (EBNA1) coding region.  
 CC A novel, claimed nucleic acid sequence which supports episomal  
 CC replication in a mammalian cell comprises an Epstein-Barr virus  
 CC (EBV) Orip sequence (see AAX03477) and an EBNA1 sequence operably  
 CC linked to a promoter, the whole sequence having a length of less  
 CC than 3 kb, and preferably less than 1.8 kb (see AAX03475). The  
 CC nucleic acid sequence preferably further comprises a bidirectional  
 CC polyadenylation consensus (see AAX03476), and is designed to be  
 CC inserted upstream of such a polyadenylation sequence in an  
 CC appropriate vector. The claimed nucleic acid sequence acts as a  
 CC compact EBV replicon. Plasmid vectors based on this sequence  
 CC replicate as episomes in the nucleus of transfected mammalian  
 CC cells. Compact EBV replicons find use in gene therapy vectors,  
 CC e.g. in gene delivery vehicles such as expression vectors.  
 CC (NB, the sequence of EBNA1 was created with respect to Claim 4 and

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Rur on: May 18, 2002, 14:59:29 ; Search time 2878.96 Seconds  
(without alignments)  
218.064 Million cell updates/sec

Title: US-09-623-329-26

Perfect score: 30

Sequence: 1 ctggtttaactggccaggaggagca 30

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank:

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_jm.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vl.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
-----					

1	30	100.0	30	6	AX018467
2	30	100.0	172281	14	EBV
3	30	100.0	184113	14	HS4B958RAJ
4	25.8	86.0	171096	14	AY037858
5	22	73.3	2661	5	AF038490
6	21.6	72.0	164086	2	AC097929
7	21.2	70.7	124124	9	AC008486
8	21.2	70.7	172188	9	AC068660
9	21	70.0	149384	9	AC006318
10	21	70.0	182045	2	AL671969
11	20.8	69.3	179310	9	AC006160
12	20.6	68.7	172486	2	AC099753
13	20.6	68.7	182196	2	AP002773
14	20.6	68.7	194281	2	AC078812
15	20.6	68.7	202833	9	AP002514
16	20.6	68.7	214498	2	AC067807
17	20.4	68.0	1755	5	AF038498
18	20.4	68.0	2705	5	AF038492
19	20.4	68.0	2715	5	AF081862
20	20.4	68.0	201268	2	AC091623
21	20.4	68.0	210869	9	CNS01DRE
22	20.2	67.3	194350	10	AC079832
23	20.2	67.3	206117	2	AC079845
24	20	66.7	1425	9	AF060981
25	20	66.7	1444	9	BC014465
26	20	66.7	40203	9	HSN104C4
27	20	66.7	63987	2	AC080154
28	20	66.7	74448	9	AL355482
29	20	66.7	140161	2	AL390725
30	20	66.7	150917	2	AC068916
31	20	66.7	152044	2	AC108103
32	20	66.7	15272	2	AC022965
33	20	66.7	161128	2	AC106755
34	20	66.7	179526	2	AC034133
35	20	66.7	180727	2	AC027221
36	20	66.7	181386	9	AC007686
37	20	66.7	181677	2	AC095165
38	20	66.7	182891	2	AC025194
39	20	66.7	189287	2	AC068362
40	20	66.7	200918	9	AC012360
41	19.8	66.0	58362	2	AC098984
42	19.8	66.0	63671	2	AC100417
43	19.8	66.0	122176	2	AL358233
44	19.8	66.0	156481	2	AL392107
45	19.8	66.0	159204	9	AL133353

## ALIGNMENTS

RESULT 1

AX018467

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BASE COUNT

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

AX018467 Sequence 26 from Patent WO9945155. 30 bp DNA linear PAT 07-SEP-2000

ORIGIN		JOURNAL MEDLINE PUBMED	10872327	8 (bases 142687 to 159853)	Bankier,A.T., Deininger,P.L., Farrell,P.J. and Barrell,B.G. Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8 Epstein-Barr virus
Query Match	100.0%	Score 30;	DB 6;	Length 30;	
Best Local Similarity	100.0%	Pred. No. 0.0029;			
Matches 30;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	ctggtttaaactggccagagagagca 30			Epstein-Barr virus
Db	1	CTGTTAAACTGGCCAGGAGGAGCA 30			Molecular biology & medicine. 1 (1), 21-45 (1983)
RESULT	2				
EBV					
LOCUS		EBV	172281 bp	DNA	circular VRL 20-SEP-1999
DEFINITION		Epstein-Barr virus (EBV) genome, strain B95-8.			
ACCESSION		V01555 J02070 K01729 K01730 V01554 X00498 X00499 X00784			
VERSION		V01555.1 GI:59074			
KEYWORDS		DNA polymerase; EBNA; genome; ribonucleotide reductase; tandem repeat; terminal repeat.			
SOURCE		Epstein-Barr virus 4			
ORGANISM		Human herpesvirus 4			
		Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gamaherpesvirinae; Lymphocryptovirus.			
REFERENCE		1 (bases 1 to 172281)			
AUTHORS		Arrand,J.R., Rymo,L., Walsh,J.E., Bjorck,E., Lindahl,T. and Griffin,B.E.			
TITLE		Molecular cloning of the complete Epstein-Barr virus genome as a set of overlapping restriction endonuclease fragments			
JOURNAL		Nucleic acids research. 9 (13), 2999-3014 (1981)			
MEDLINE		82014887			
PUBMED		6269068			
REFERENCE		2 (bases 1 to 172281)			
AUTHORS		Kozak,M.			
TITLE		Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes			
JOURNAL		Nucleic acids research. 9 (20), 5233-5262 (1981)			
MEDLINE		82059504			
PUBMED		7301588			
REFERENCE		3 (bases 1 to 172281)			
AUTHORS		Deininger,P.L., Bankier,A., Farrell,P., Baer,R. and Barrell,B.			
TITLE		Sequence analysis and in vitro transcription of portions of the Epstein-Barr virus genome			
JOURNAL		Journal of cellular biochemistry. 19 (3), 267-274 (1982)			
MEDLINE		83109311			
PUBMED		6296170			
REFERENCE		4 (bases 1 to 172281)			
AUTHORS		Farrell,P.J., Deininger,P.L., Bankier,A. and Barrell,B.			
TITLE		Homologous upstream sequences near Epstein-Barr virus promoters			
JOURNAL		Proceedings of the National Academy of Sciences of the United States of America. 80 (6), 1565-1569 (1983)			
MEDLINE		83169725			
PUBMED		6300857			
REFERENCE		5 (bases 45644 to 52450)			
AUTHORS		Jeang,K.T. and Hayward,S.D.			
TITLE		Organization of the Epstein-Barr virus DNA molecule. III. Location of the P3HR-1 deletion junction and characterization of the NotI repeat units that form part of the template for an abundant 12-O-tetradecanoylphorbol-13-acetate-induced mRNA transcript			
JOURNAL		Journal of virology. 48 (1), 135-148 (1983)			
MEDLINE		83294686			
PUBMED		8310141			
REFERENCE		6 (bases 159853 to 172281)			
AUTHORS		Bankier,A.T., Deininger,P.L., Satchwell,S.C., Baer,R., Farrell,P.J. and Barrell,B.G.			
TITLE		DNA sequence analysis of the EcoRI Dhet fragment of B95-8 Epstein-Barr virus containing the terminal repeat sequences			
JOURNAL		Molecular biology & medicine. 1 (4), 425-445 (1983)			
MEDLINE		85060428			
PUBMED		6094955			
REFERENCE		7 (bases 1 to 172281)			
AUTHORS		Farrell,P.J., Bankier,A., Sequin,C., Deininger,P. and Barrell,B.G.			
TITLE		Latent and lytic cycle promoters of Epstein-Barr virus			
JOURNAL		The EMBO journal. 2 (8), 1331-1338 (1983)			
MEDLINE		20331131			
PUBMED		10872327			
REFERENCE		8 (bases 142687 to 159853)			
AUTHORS		Bankier,A.T., Deininger,P.L., Farrell,P.J. and Barrell,B.G.			
TITLE		Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8 Epstein-Barr virus			
JOURNAL		Molecular biology & medicine. 1 (1), 21-45 (1983)			
MEDLINE		85035713			
PUBMED		6092825			
REFERENCE		9 (bases 112620 to 125316)			
AUTHORS		Sequin,C., Farrell,P.J. and Barrell,B.G.			
TITLE		DNA sequence and transcription of the BamHI fragment B region of B95-8 Epstein-Barr virus			
JOURNAL		Molecular biology & medicine. 1 (3), 369-392 (1983)			
MEDLINE		85060424			
PUBMED		6094953			
REFERENCE		10 (bases 45415 to 52824)			
AUTHORS		Jones,M.D., Foster,L., Sheedy,T. and Griffin,B.E.			
TITLE		The EB virus genome in Daudi Burkitt's lymphoma cells has a deletion similar to that observed in a non-transforming strain (P3HR-1) of the virus			
JOURNAL		The EMBO journal. 3 (4), 813-821 (1984)			
MEDLINE		84207939			
PUBMED		6327290			
REFERENCE		11 (bases 87650 to 92703)			
AUTHORS		Biggin,M., Farrell,P.J. and Barrell,B.G.			
TITLE		Transcription and DNA sequence of the BamHI L fragment of B95-8 Epstein-Barr virus			
JOURNAL		The EMBO journal. 3 (5), 1083-1090 (1984)			
MEDLINE		84236104			
PUBMED		6203743			
REFERENCE		12 (bases 7315 to 9312)			
AUTHORS		Yates,J., Warren,N., Reisman,D. and Sugden,B.			
TITLE		A cis-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells			
JOURNAL		Proceedings of the National Academy of Sciences of the United States of America. 81 (12), 3806-3810 (1984)			
MEDLINE		84222045			
PUBMED		6328526			
REFERENCE		13 (bases 76089 to 79808)			
AUTHORS		Gibson,T., Stockwell,P., Ginsburg,M. and Barrell,B.			
TITLE		Homology between two EBV early genes and HSV ribonucleotide reductase and 38K genes			
JOURNAL		Nucleic acids research. 12 (12), 5087-5099 (1984)			
MEDLINE		84247360			
PUBMED		6330697			
REFERENCE		14 (bases 1 to 172281)			
AUTHORS		Baer,R.J., Bankier,A.T., Biggin,M.D., Deininger,P.L., Farrell,P.J., Gibson,T.J., Hatfull,G.F., Hudson,G.S., Satchwell,S.C., Sequin,C., Tuffnell,P.S. and Barrell,B.G.			
TITLE		DNA sequence and expression of the B95-8 Epstein-Barr virus genome			
JOURNAL		Nature. 310 (5974), 207-211 (1984)			
MEDLINE		84270667			
PUBMED		6087149			
REFERENCE		15 (bases 1 to 172281)			
AUTHORS		Bodescu,M. and Perricaudet,M.			
TITLE		Cloned alternative splice sites in Epstein-Barr virus RNAs			
JOURNAL		Nucleic acids research. 15 (14), 5887 (1987)			
MEDLINE		87289053			
PUBMED		3039467			
REFERENCE		16 (bases 1 to 172281)			
AUTHORS		Laux,G., Perricaudet,M. and Farrell,P.J.			
TITLE		A spliced Epstein-Barr virus gene expressed in immortalized lymphocytes is created by circularization of the linear viral genome			
JOURNAL		The EMBO journal. 7 (3), 769-774 (1988)			
MEDLINE		88283646			
PUBMED		2840285			
REFERENCE		17 (bases 1 to 172281)			
AUTHORS		Hatfull,G.F., Barrell,B.G., Quinlan,J. and McGeoch,D.			
JOURNAL		Unpublished			



REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

18 (bases 1 to 172281)  
Farrell, P.J. and Barrell, B.G.  
Direct Submission  
Submitted (05-JUN-1984)  
19 (bases 1 to 172281)  
Farrell, P.J.  
Direct Submission  
Submitted (18-MAR-1988) Farrell P., Ludwig Institute for Cancer  
Research, St. Mary's Hospital Medical School, Norfolk Place London  
W2 1PG

## COMMENT

Listed under this feature are all known protein coding regions as well as all the major open reading frames in the sequence. In general the term major is taken as the longest frame in a particular region taking into account the adjacent longest frames and likely transcription signals. Note that on this basis some long overlapping frames have been excluded and on the other hand some small frames have been included which might represent exons or genes because they occur in a logical combination with other features or because of some other experimental data. The reading frames are named according to the Bam HI fragment in which they start. eg BAF3 is the third leftward frame starting in Bam HI fragment A. B0RF1 is the first rightward frame in Bam HI fragment O. If there is an obvious TATA sequence followed by an in frame Met codon that satisfies the rules of Kozak [12] in that there is a purine at -3 and/or a G at +4 then the reading frame is numbered from the A of the ATG to the base preceding the termination codon. If there is no obvious initiation codon or there is a substantial reading frame in phase before the ATG then the reading frame is numbered from the first base of the first codon.

SITES OF POLYA signals  
This feature lists all occurrences of the sequence AATAAA which is found normally approximately 20 bases upstream of the mRNA processing/polyA addition site. The rarely used homolog AITAAA is only listed when it is found in a position close to the end of a major reading frame.

SITES OF DONOR and ACCEPT sequences  
This is not a comprehensive listing of all such sequences and only the positions of a few have been noted because they occur in potentially interesting positions. The number quoted in the table is the position of the terminal base in the intron in each case. Restriction enzyme SITES.  
Only the positions of the sites Bam HI (BAM) are listed.

## RPT

This feature is used to define repetitive sequences.

SITE DEL  
This feature defines deletions in B95-8 with respect to other strains such as RAJ1 and also to deletions in other strains such as PJHRI and DAUDI with respect to B95-8.

## SITE HPN

Denotes sequences with twofold symmetry in could form hairpin loops. This is not a comprehensive list - only a few occurrences noted.

## OKGRPL

Denotes the region that encompasses an origin of replication (ori P). [13].

## NUMBERING

The DNA sequence of B95-8 EBV has been revised [19]. The original (Baer et al, 1984) base 359 has been deleted so the new sequence around that position reads TCAGTCTTT. To avoid renumbering the entire sequence, position 1 has been moved 1 base to the left of the EcoRI site separating EcoRI Dhet from EcoRI I (ie the first A of AGAATTC).

## FEATURES

source  
1. .172281  
/organism="Human herpesvirus 4"  
/strain="B95-8"  
/db\_xref="taxon:10376"

Query Match 100.0%; Score 30; DB 14; Length 172281;  
Best Local Similarity 100.0%; Pred. NO. 0.0021;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctggttttaaacaggccaggagagagca 30  
|||||  
Db 165644 CTGGTTTAAACTGGCCAGGAGAGGCA 165673

## RESULT 3

HS4B958RAJ  
LOCUS

## DEFINITION

Epstein-Barr virus, artificial joining of B95-8 complete genome and the sequences from Raji of the large deletion found in B95-8.

## ACCESSION

M80517 M75989

## KEYWORDS

Human herpesvirus 4 DNA.

## SOURCE

Human herpesvirus 4

## ORGANISM

Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae; Lymphocryptovirus.

## REFERENCE

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

1 (sites)

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGGTTAACTGGCCAGAGAGCA 30  
|||||  
b 177475 CTGGTTAACTGGCCAGAGAGCA 177504

RESULT 4  
AY037858 171096 bp DNA circular VRL 31-DEC-2001  
LOCUS  
DEFINITION Cercopithecine herpesvirus 15 strain LCL8664, complete genome.  
ACCESSION AF037858 AF148640 AF148641 AF159308 AF159309 AF159310  
AF227123 AF227124 AF25963 093160 093909  
VERSION AY037858.1 GI:18025465  
KEYWORDS  
SOURCE cercopithecine herpesvirus 15.  
ORGANISM cercopithecine herpesvirus 15.  
Viruses: dsDNA viruses, no RNA stage; Herpesviridae;  
Gammaherpesvirinae; Lymphocryptovirus.  
Wang, F.  
REFERENCE  
AUTHORS Franken, M., Devergne, O., Rosenzweig, M., Annis, B., Kieff, E. and  
Wang, F.  
TITLE Comparative analysis identifies conserved tumor necrosis factor  
receptor-associated factor 3 binding sites in the human and simian  
Epstein-Barr virus oncogene LMP1  
J. Virol. 70 (11), 7819-7826 (1996)  
97048062  
PUBMED 8892903  
REFERENCE 2 (bases 1 to 171096)  
AUTHORS Rivallier, P., Quink, C. and Wang, F.  
TITLE Strong selective pressure for evolution of an Epstein-Barr virus  
LMP2B homologue in the rhesus lymphocryptovirus  
J. Virol. 73 (10), 8867-8872 (1999)  
99412410  
PUBMED 10482645  
REFERENCE 3 (bases 1 to 171096)  
AUTHORS Jiang, H., Cho, Y.-G. and Wang, F.  
TITLE Structural, functional, and genetic comparisons of Epstein-Barr  
virus nuclear antigen 3A, 3B, and 3C homologues encoded by the  
rhesus lymphocryptovirus  
J. Virol. 74 (13), 5921-5932 (2000)  
20304984  
PUBMED 10970361  
REFERENCE 4 (bases 1 to 171096)  
AUTHORS Rao, P., Jiang, H. and Wang, F.  
TITLE Cloning of the rhesus lymphocryptovirus viral capsid antigen and  
Epstein-Barr virus-encoded small RNA homologues and use in  
diagnosis of acute and persistent infections  
J. Clin. Microbiol. 38 (9), 3219-3225 (2000)  
20440633  
PUBMED 10970361  
REFERENCE 5 (bases 1 to 171096)  
AUTHORS Rivallier, P., Jiang, H., Cho, Y.-G., Quink, C. and Wang, F.  
TITLE Complete Nucleotide Sequence of the Rhesus Lymphocryptovirus:  
Genetic Validation for an Epstein-Barr Virus Animal Model  
J. Virol. 76 (1), 421-426 (2002)  
11739708  
PUBMED 11739708  
REFERENCE 6 (bases 1 to 171096)  
AUTHORS Wang, F., Franken, M. and Annis, B.  
TITLE Direct Submission  
Submitted (11-JAN-1996) Medicine, Brigham and Women's Hospital, 181  
Longwood Avenue, Boston, MA 02115, USA  
7 (bases 1 to 171096)  
REFERENCE 7 (bases 1 to 171096)  
AUTHORS Moghaddam, A., Koch, J., Annis, B. and Wang, F.  
TITLE Direct Submission  
Submitted (12-MAR-1997) Medicine, Brigham and Women's Hospital, 181  
Longwood Avenue, Boston, MA 02115, USA  
8 (bases 1 to 171096)  
REFERENCE 8 (bases 1 to 171096)  
AUTHORS Moghaddam, A., Annis, B. and Wang, F.  
TITLE Direct Submission  
Submitted (17-MAR-1997) Medicine, Brigham and Women's Hospital, 181  
Longwood Avenue, Boston, MA 02115, USA  
9 (bases 1 to 171096)  
REFERENCE 9 (bases 1 to 171096)

AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REMARK  
COMMENT  
FEATURES  
Source

CDS  
CDS  
misc\_feature

1. 171096  
/organism="cercopithecine herpesvirus 15"  
/strain="LCL8664"  
/db\_xref="taxon:104228"  
join(141..354,437..535,644..892,974..1054,1134..1289,  
1377..1592,1673..1779)  
/note="similar to Epstein-Barr virus LMP2B"  
/codon\_start=1  
/product="LMP2B"  
/protein\_id="AAK95410.1"  
/db\_xref="GI:18025466"  
/translation="MAEGLPVVSAPYLFWLAGIAAFCFASVSVLVVTTGLASLLLL  
LAALVSYAQRRAKAVTKLTLVGIIVFFAILLTWVQPKDAIVFALLTGAALVQA  
IYILVLLFLPTVRRRKMIVCVVLLFACAVLLVDAIWLSPGLSMTVVALFL  
LLAFIWLSSPMGIGGAAALLLAAALGLSLVGLGELNATMFLMFLVLTLLV  
LSPFPTQMYLMSHLLYALSLTASALLAGSILLOTWGGDNTFFPFLSCLMLLIV  
AGILFALITWGTGSKTYCPVFCVLSGLLTMAGLVVWVLMIKLLSAMLITAGCF  
IFELGFIPLGLKFCRLCCCLKLESGDQVATYQNNV"  
1836..5780  
/note="similar to Epstein-Barr virus BNRF1"  
/codon\_start=1  
/product="BNRF1"  
/protein\_id="AAK95411.1"  
/db\_xref="GI:18025467"  
/translation="MAGSRVPPQAHGSOFLMIRLFGTDGNGVUEQRLYELLSNPRSE  
LGLSTGPLVSENLLTVAILRGNDPPRQPERARELALVGLIILGNSQGERUGTESAL  
NALNKYVYAYGPDMMARPTWSAEIQOFLRLGLTKYVLRVGLGRQGFVETGARPSE  
RQFOAI NIHVLFDNLVRKYPGQASAGFQALLVPPTVDVPRDLRTVNEWVFGGVA  
GGHOLADELMI VSPLRDYSGLHVLPTDTEALDVMRASNREAREATHSEHG FVHTACT  
IOPNCPOLFMHROHPLFPFVSATASSLWGYYPALAGSPDVREKARROCAFOTKAVA  
ECHARSGVPVAGRYRTVDTLRGEGCLTPTMETGELGVIHQHAI DTRARVEYGOYLVN  
LGPFPNSLTPAPCPYAESALAGATVQTALFLPSGLYPAPCVSSHARPGPCALIEH  
LRSLVQKGGLLFLSHLPEDVKGGIDGTAQATSRGIKOTVETIFLNPNCNNFLIV  
KORGEKINGVQLQALKEACVAGCQYVILGTVPLGLNPNVNDLSPVSAETMDDF  
APFESVEFPSTPEGEMSPVHSDTEEMDYSPPSHSSWLSLSCSLTSLVSHPSVGSKE  
HLVHRTDRVSGGRVALQGVGLDPLADYAFVAHSQVWVRPGCAPPLPYRTWDRME  
KLLVSARPTDEAPKVSCTVITLGEGYKVTLDPIETGPKLAWVEALMNAACAPLVNPD  
VILLNLHADPGRVDTAVMDLMTAASDYARDLGVKLTYGSACPETGAGTGMVTV  
VTASARFSAPLITPVLPQRSGLLSVRRGDKIQGSLFEQVLFSDVATAPRAPQAD  
AMKTI FRTVOGLKGLVLSGHDVSDGGLVTCVLEMALAGORGVTISVPAADALPOL  
LSEHPGLVFEVEGNMTEVMOTLRQNIYPAVLRGVGEAGADQFVQHGSEVTFROS  
FTSLLSWTSFTATQYEFRLRGLNRNSHVSYDGENEALAI SPLTKNKLRLVTFEP  
DPRCAVAVLCAPGTRGHSLLAATNAGCLCQRYFRDVRRTFFDKYAGLAVGVGVA  
ADAPLAGATVALIDRSPLADALLAFARPETFSLGELGVLQVLAGVAGSTDP  
TPEPCVENVQBSPLILTNNASGLFESRWLNFSPHDTKSVMLRGIRGCVLPQWIGS  
HLGELFTNPAMPYVQLNQRHQTAFHSHRTDPRRHAEHYPRNPSERSNIGLCSPDOR  
HLAULCDFSLPTDFWQWEYIPSAFAPYTGSPFLMFOAHLAKSRHSPSS"  
6653..6824  
/note="similar to Epstein-Barr virus EBER-1"

misc\_feature  
6653..6824  
/note="similar to Epstein-Barr virus EBER-1"



Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carrroll, L., Dederich, D.A., Delaney, K.K., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escoto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hui, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratoch, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtharge, O., Lieu, C., Liu, J., Liu, W., Louisgied, H., Lozano, R., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogulu, M., Okunolu, G., Oragunye, N., Owiedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojebokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sigson, I., Sodergren, E., Sonalike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalobos, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Woodson, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

Direct Submission  
Unpublished  
2. (bases 1 to 164086)  
Worley, K.C.

Direct Submission  
Submitted (23-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Dec 20, 2001 this sequence version replaced gi:16327635.

----- Genome Center -----  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information -----  
Center project name: GFOL  
Center clone name: CH230-76K21  
----- Summary Statistics -----  
Assembly program: Phrap; version 0.990429first call to findPhrapList  
Consensus quality: 134053 bases at least Q40  
Consensus quality: 145317 bases at least Q30  
Consensus quality: 154273 bases at least Q20  
Estimated insert size: 142483; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; aga.ose-ip estimation  
Quality coverage: 2.1x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 68 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 9231: contig of 9231 bp in length  
\* 9232 9331: gap of unknown length  
\* 9332 15115: contig of 5784 bp in length

15116  
15216  
15216  
19815: gap of unknown length  
19816  
22272: gap of unknown length  
22273  
22273  
26953: gap of unknown length  
27053  
32578: gap of unknown length  
32579  
32579  
35079: gap of unknown length  
35080  
35179: gap of unknown length  
39903  
40003: gap of unknown length  
40004  
44669: gap of unknown length  
44670  
49120: gap of unknown length  
49121  
51996: gap of unknown length  
52096: gap of unknown length  
57220: contig of 5124 bp in length  
57220  
57320: gap of unknown length  
60946: gap of unknown length  
61046  
61046  
63911: contig of 2865 bp in length  
64011: gap of unknown length  
64011  
67167: contig of 3156 bp in length  
67267: gap of unknown length  
69428: contig of 2161 bp in length  
69428  
69528: gap of unknown length  
72109: contig of 2481 bp in length  
72109  
75099: gap of unknown length  
75099  
75476: contig of 3367 bp in length  
75476  
75576: gap of unknown length  
75576  
78502: contig of 2926 bp in length  
78502  
78603  
81112: contig of 2510 bp in length  
81112  
83320: contig of 2108 bp in length  
83320  
83420: gap of unknown length  
85132: contig of 1712 bp in length  
85132  
85232: gap of unknown length  
85232  
86945: contig of 1713 bp in length  
86945  
87045: gap of unknown length  
87045  
89815: contig of 2770 bp in length  
89815  
89915: gap of unknown length  
92131: contig of 2216 bp in length  
92131  
92231: gap of unknown length  
94673: contig of 2442 bp in length  
94673  
94773: gap of unknown length  
96243: contig of 1469 bp in length  
96243  
96342: gap of unknown length  
96342  
98599: gap of unknown length  
98599  
101707: contig of 3008 bp in length  
101707  
101807: gap of unknown length  
103334: contig of 1527 bp in length  
103334  
103434: gap of unknown length  
103434  
105425: contig of 1991 bp in length  
105425  
105523: gap of unknown length  
105523  
107552: contig of 2027 bp in length  
107552  
107552: gap of unknown length  
109348: contig of 1696 bp in length  
109348  
109448: gap of unknown length  
111256: contig of 1808 bp in length  
111256  
111356: gap of unknown length  
111356  
112795: contig of 1439 bp in length  
112795  
112895: gap of unknown length  
112895  
115009: contig of 2114 bp in length  
115009  
115109: gap of unknown length  
115109  
116786: contig of 1677 bp in length  
116786  
116886: gap of unknown length  
116886  
118516: contig of 1630 bp in length  
118516  
118616: gap of unknown length  
118616

*	118617	121061:	contig	of	2445	bp	in	length
*	121062	121161:	gap	of	unknown	length		
*	121162	122204:	contig	of	1043	bp	in	length
*	122205	122304:	gap	of	unknown	length		
*	122305	123964:	contig	of	1660	bp	in	length
*	123965	124064:	gap	of	unknown	length		
*	124065	125463:	contig	of	1399	bp	in	length
*	125464	125563:	gap	of	unknown	length		
*	125564	126633:	contig	of	1072	bp	in	length
*	126636	126735:	gap	of	unknown	length		
*	126736	128662:	contig	of	1927	bp	in	length
*	128663	128762:	gap	of	unknown	length		
*	128763	130538:	contig	of	1776	bp	in	length
*	130539	130638:	gap	of	unknown	length		
*	130639	132068:	contig	of	1430	bp	in	length
*	132069	132168:	gap	of	unknown	length		
*	132169	133386:	contig	of	1218	bp	in	length
*	133387	133486:	gap	of	unknown	length		
*	133487	135330:	contig	of	1844	bp	in	length
*	135331	135430:	gap	of	unknown	length		
*	135431	137222:	contig	of	1792	bp	in	length
*	137223	137322:	gap	of	unknown	length		
*	137323	138672:	contig	of	1350	bp	in	length
*	138673	138772:	gap	of	unknown	length		
*	138773	140173:	contig	of	1401	bp	in	length
*	140174	140273:	gap	of	unknown	length		
*	140274	141598:	contig	of	1325	bp	in	length
*	141599	141698:	gap	of	unknown	length		
*	141699	143484:	contig	of	1786	bp	in	length
*	143485	143584:	gap	of	unknown	length		
*	143585	144851:	contig	of	1267	bp	in	length
*	144852	144951:	gap	of	unknown	length		
*	144952	146120:	contig	of	1169	bp	in	length
*	146121	146220:	gap	of	unknown	length		
*	146221	147798:	contig	of	1578	bp	in	length
*	147799	147898:	gap	of	unknown	length		

Query Match	72.0%	Score	21.6;	DB	2;	Length	164086;
Best Local Similarity	85.7%;	Pred. No.	18;				
Matches	24;	Conservative	0;	Mismatches	4;	Indels	0;
						Gaps	0;
QY	3	ggtttaaactggccaggagagagca	30				
Db	131040	GGGTTAACTGGGCCAGGAGAGAGNA	131067				
RESULT	7						
LOCUS	AC008486						
DEFINITION	Homo sapiens chromosome 5 clone CTC-421B9, complete sequence.	DNA	124124 bp	linear	PRI 18-MAY-2001		
ACCESSION	AC008486						
VERSION	AC008486..5	GI:	14141751				
KEYWORDS	HTC.						
SOURCE	human.						
ORGANISM	Homo sapiens						
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.						
REFERENCE	1	(bases 1 to 124124)					
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.						
TITLE	Direct Submission						
JOURNAL	Unpublished						
REFERENCE	2	(bases 1 to 124124)					
AUTHORS	DOE Joint Genome Institute.						
TITLE	Direct Submission						
JOURNAL	Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA						
REFERENCE	3	(bases 1 to 124124)					
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.						
TITLE	Direct Submission						
JOURNAL	Submitted (18-MAY-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA						
COMMENT	On May 18, 2001 this sequence version replaced gi:13699364. Draft Sequence Produced by DOE Joint Genome Institute						

www.jni.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap quality >=40 99.8% of Sequence:  
Estimated Total Number of Errors is 0.2.  
STS Content:  
SHGC-7529 G16911  
SHGC-140030 G62459  
WI-5164 G04780.

FEATURES	SOURCE
----------	--------

BASE COUNT	ORIGIN
38425 a	23117 c 24743 g 37839 t

```

Query Match          70.7%; Score 21.2; DB 9; Length 124124;
Best Local Similarity 88.5%; Pred. No. 28;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 ctggtttaaactgggccagagagagg 26
      || |||| ||||| ||||| |||||
Db 118346 CTTGTTTCACTGGACCCAGGAGAGG 118371

RESULT      8
AC068660/c 172188 bp DNA linear PRI 28-FEB-2001
LOCUS      AC068660
DEFINITION Homo sapiens chromosome 5 clone CTC-349J2, complete sequence.

```

FEATURES source

BASE COUNT	ORIGIN
55260 a 33373 c 31527 g	52028 t

Query Match 70.7%; Score 21.2; DB 9; Length 172188;  
 Best Local Similarity 88.5%; Pred. No. 28;  
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ctgggttaactggccaggagagg 26  
 Db 54736 CTTGTTTCACTGGACCCGAGAGG 54711

RESULT 9  
 LOCUS AC006318 149384 bp DNA linear PRI 25-MAR-2001  
 DEFINITION Homo sapiens clone RP4-728119, complete sequence.  
 ACCESSION AC006318  
 VERSION AC006318.2 GI:5757549  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 149384)  
 AUTHORS Waterston,R.H.  
 TITLE The sequence of Homo sapiens clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 149384)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-JAN-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 3 (bases 1 to 149384)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-AUG-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 4 (bases 1 to 149384)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 REFERENCE 5 (bases 1 to 149384)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-MAR-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 COMMENT On Aug 21, 1999 this sequence version replaced gi:4204348.  
 Center project name: H\_DJ0728119.  
 FEATURES  
 source  
 1..149384  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="RP4-728119"  
 BASE COUNT 47251 a 32070 c 28867 g 41196 t

Query Match 70.0%; Score 21; DB 9; Length 149384;  
 Best Local Similarity 82.8%; Pred. No. 34;  
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 2 tgggttaactggccaggagagagca 30  
 Db 32846 TGGTTTAACTGGCCTTGGATGGCAGCA 32874

RESULT 10  
 LOCUS AL671969/c 182045 bp DNA linear HTG 08-FEB-2002  
 DEFINITION Mus musculus chromosome 11 clone RP23-33907, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, in unordered pieces.

AL671969  
 AL671969.1 GI:18643886  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 house mouse.  
 SOURCE Mus musculus  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (sites)  
 REFERENCE Burton,J.  
 AUTHORS Direct Submission  
 TITLE Submitted (07-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,  
 JOURNAL Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 COMMENT  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquery@sanger.ac.uk  
 ----- Project Information  
 Center project name: BM33907  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Consensus quality: 179231 bases at least Q40  
 Consensus quality: 179936 bases at least Q30  
 Consensus quality: 180435 bases at least Q20  
 Insert size: 181045; sum-of-contigs  
 Insert size: 189654; 12.2% error; agarose-fp  
 Quality coverage: 8.23x in Q20 bases; sum-of-contigs Quality  
 coverage: 8.26x in Q20 bases; agarose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved  
 FEATURES  
 source  
 1..182045  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="11"  
 /clone="RP23-33907"  
 /clone\_lib="RPCI-23"  
 misc\_feature 1..8321  
 /note="assembly\_fragment:03520  
 fragment\_chain:1  
 clone\_end:SP6  
 vector\_side:left"  
 8422..15230  
 /note="assembly\_fragment:03808  
 fragment\_chain:1"  
 misc\_feature 15331..20621  
 /note="assembly\_fragment:00624"  
 20722..29561  
 /note="assembly\_fragment:01536  
 fragment\_chain:2"  
 misc\_feature 29662..42290  
 /note="assembly\_fragment:03510  
 fragment\_chain:2"  
 misc\_feature 42391..63901  
 /note="assembly\_fragment:03666  
 fragment\_chain:2"  
 misc\_feature 64002..77755  
 /note="assembly\_fragment:00556  
 fragment\_chain:2"  
 misc\_feature 77856..80051  
 /note="assembly\_fragment:03278  
 fragment\_chain:2"  
 misc\_feature 80152..135084  
 /note="assembly\_fragment:00109  
 fragment\_chain:2"  
 misc\_feature 135185..176750  
 /note="assembly\_fragment:02455  
 fragment\_chain:2"

misc\_feature 176851..182045  
/note="assembly\_fragment:01655  
fragment\_chain:2  
clone\_end:T7  
vector\_side:right"  
BASE COUNT 49816 a 38846 c 39841 g 52536 t 1006 others  
ORIGIN

Query Match 70.0%; Score 21; DB 2; Length 182045;  
Best Local Similarity 82.8%; Pred. No. 34;  
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ctggtttaactggccaggagagc 29  
||||| ||||| ||||| ||||| |||||  
Db 109274 CTGATTTTAACTCAGCCAGGAGCAGC 109246

RESULT 11  
AC006160/c  
LOCUS  
DEFINITION Homo sapiens chromosome 4 clone C0481P14 map 4p16, complete  
sequence.  
ACCESSION AC006160  
VERSION AC006160.9 GI:5701616  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 179310)  
AUTHORS Stone,N.E., Schmutz,J.J., Shang,J., Pennacchio,L.A., Cox,D.R. and  
Myers,R.M.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 179310)  
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.  
TITLE Direct Submission  
JOURNAL Submitted (08-DEC-1998) Department of Genetics, Stanford Human  
Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA  
REFERENCE 3 (bases 1 to 179310)  
AUTHORS Stone,N.E., Schmutz,J.J., Shang,J., Pennacchio,L.A., Cox,D.R. and  
Myers,R.M.  
TITLE Direct Submission  
JOURNAL Submitted (04-AUG-1999) Department of Genetics, Stanford Human  
Genome Center, 4005 Miranda Avenue, Palo Alto, CA 94304, USA  
COMMENT On Aug 4, 1999 this sequence version replaced gi:4580748.  
Quality: Phrap Quality >=40 99.9% of Sequence;  
Estimated Total Number of Errors is 0.1.  
STS Content:  
SHGC-36184 G30090  
WI-13901 G21191  
SHGC-50339 G33480  
SHGC-50545 G33619  
WI-15951 G24340  
SHGC-79088 G48789.  
Location/Qualifiers  
1..179310  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4p16"  
/clone="C0481P14"  
/clone\_lib="ROSWELL PARK CANCER RPCI - 11 Human Male BAC  
Library"

BASE COUNT 50020 a 40057 c 39508 g 49725 t  
ORIGIN

Query Match 69.3%; Score 20.8; DB 9; Length 179310;  
Best Local Similarity 91.7%; Pred. No. 42;  
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ctggtttaactggccaggagagc 24  
||||| ||||| ||||| ||||| |||||  
Db 137849 CTGTTTCCACTGGCCAGGAGA 137826

RESULT 12  
AC099753/c  
LOCUS  
DEFINITION Homo sapiens chromosome 3 clone RP11-466A13, WORKING DRAFT  
SEQUENCE, 4 unordered pieces.  
ACCESSION AC099753  
VERSION AC099753.1 GI:17017611  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 172486)  
AUTHORS Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 172486)  
AUTHORS Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.  
TITLE Direct Submission  
JOURNAL Submitted (20-NOV-2001) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
COMMENT On Nov 20, 2001 this sequence version replaced gi:8078187.  
----- Genome Center  
Center: University of Washington Genome Center  
Center Code: UWGC  
Web site: http://www.genome.washington.edu  
Contact: uwgctgs@u.washington.edu  
Drafting Center: BCM  
----- Project Information  
Center project name: chr-3  
Center clone name: RP11-466A13 (bc0435)  
----- Summary Statistics  
Sequencing vector: unknown; 64% of reads  
Sequencing vector: plasmid; L08752; 36% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 169490 bases at least Q40  
Consensus quality: 171107 bases at least Q30  
Consensus quality: 171810 bases at least Q20  
Insert size: 175800; 9.1% error; agarose-fp  
Insert size: 172186; sum-of-contigs  
Quality coverage: 6.2x in Q20 bases; agarose-fp  
Quality coverage: 6.3x in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 19647: contig of 19647 bp in length  
\* 19648 19747: gap of unknown length  
\* 19748 58761: contig of 39014 bp in length  
\* 58762 58861: gap of unknown length  
\* 58862 108948: contig of 50087 bp in length  
\* 108949 109048: gap of unknown length.  
\* 109049 172486: contig of 63438 bp in length.  
Location/Qualifiers  
1..172486  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/clone="RP11-466A13"  
/clone\_lib="RPCI human BAC library 11"  
1..19647  
/note="assembly\_name:Contig69"

FEATURES  
source  
1..172486  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/clone="RP11-466A13"  
/clone\_lib="RPCI human BAC library 11"  
1..19647  
misc\_feature  
/note="assembly\_name:Contig69"

```
misc_feature 19748..58761
              /note="assembly_name:Contig70"
misc_feature 58862..108948
              /note="assembly_name:Contig71"
misc_feature 109049..172486
              /note="assembly_name:Contig72"
BASE COUNT 55434 a 30908 c 31419 g 54425 t 300 others
ORIGIN

Query Match      68.7%; Score 20.6; DB 2; Length 172486;
Best Local Similarity 85.2%; Pred. No. 53;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 gtttaactggcccgaggagagca 30
    ||||| ||||| ||||| |||||
Db 16731 GTTTATAATGGCCAGGAGATGACGA 16705

RESULT 13
AP002773
LOCUS      182196 bp      DNA      linear      HTG 13-JUL-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-626J19 map 11q, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
ACCESSION AP002773
VERSION    1 GI:9188611
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     Homo sapiens DNA, clone:RP11-626J19.
ORGANISM   Homo sapiens

REFERENCE 1 (bases 1 to 182196)
AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE      Homo sapiens 182196 genomic DNA of 11q
JOURNAL    Published Only in Database (2000) In press
REFERENCE 2 (bases 1 to 182196)
AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE      Direct Submission
JOURNAL    Submitted (10-JUL-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gscc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-42-778-9923, Fax:81-42-778-9924)
----- Genomic Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gscc.riken.go.jp
----- Project Information
Center project name: HumDraft11
(enter clone name: RP11-626J19)
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 175904 bases at least Q40
Consensus quality: 179212 bases at least Q30
Consensus quality: 180392 bases at least Q20
Insert size: 180996; sum-of-contigs
Quality coverage: 5.92x in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently consists of
13 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved 1
35116 70079 contig of 35015 bp in length
70180 98586 contig of 34964 bp in length

misc_feature 1..35015
              /note="assembly_fragment clone_end:SP6 vector_side:right"
misc_feature 35116..70079
              /note="assembly_fragment clone_end:SP6 vector_side:right"
misc_feature 70180..98586
              /note="assembly_fragment clone_end:SP6 vector_side:right"
misc_feature 98687..121363
              /note="assembly_fragment clone_end:SP6 vector_side:right"
misc_feature 121464..138890
              /note="assembly_fragment clone_end:SP6 vector_side:right"
misc_feature 138991..152450
              /note="assembly_fragment clone_end:SP6 vector_side:right"
misc_feature 152551..165251
              /note="assembly_fragment clone_end:SP6 vector_side:right"
misc_feature 165352..169788
              /note="assembly_fragment clone_end:SP6 vector_side:right"
misc_feature 169889..173158
              /note="assembly_fragment clone_end:SP6 vector_side:right"
misc_feature 173259..176622
              /note="assembly_fragment clone_end:SP6 vector_side:right"
misc_feature 176723..179274
              /note="assembly_fragment clone_end:SP6 vector_side:right"
misc_feature 179375..180892
              /note="assembly_fragment clone_end:SP6 vector_side:right"

FEATURES
Source
1..182196
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-626J19"
1..35015
/note="assembly_fragment clone_end:T7 vector_side:left"
35116..70079
/note="assembly_fragment"
70180..98586
/note="assembly_fragment"
98687..121363
/note="assembly_fragment"
121464..138890
/note="assembly_fragment"
138991..152450
/note="assembly_fragment"
152551..165251
/note="assembly_fragment"
165352..169788
/note="assembly_fragment"
169889..173158
/note="assembly_fragment"
173259..176622
/note="assembly_fragment"
176723..179274
/note="assembly_fragment"
179375..180892
/note="assembly_fragment"
```



JOURNAL

COMMENT

Submitted (04-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Dec 29, 2000 this sequence version replaced gi:10180082.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: HBUK  
Center Clone name: Rp11-457K5  
----- Summary Statistics  
Sequencing vector: M13; L08821  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 141660 bases at least Q40  
Consensus quality: 163023 bases at least Q30  
Consensus quality: 174681 bases at least Q20  
Estimated insert size: 174878; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-1p estimation  
Quality coverage: 2.5x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 45 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 15573: contig of 15573 bp in length  
15574 15673: gap of unknown length  
15674 27775: contig of 12102 bp in length  
27775 27875: gap of unknown length  
27875 27935: contig of 10060 bp in length  
27935 38035: gap of unknown length  
38035 49300: contig of 11265 bp in length  
49300 49400: gap of unknown length  
49400 58955: contig of 9555 bp in length  
58955 59055: gap of unknown length  
59055 60622: contig of 7007 bp in length  
60622 66162: gap of unknown length  
66162 73333: contig of 7171 bp in length  
73333 79376: contig of 5943 bp in length  
79376 79476: gap of unknown length  
79476 86199: contig of 6723 bp in length  
86199 86299: gap of unknown length  
86299 92377: contig of 6077 bp in length  
92377 92477: gap of unknown length  
92477 97779: contig of 5303 bp in length  
97779 97879: gap of unknown length  
97879 104591: contig of 6712 bp in length  
104591 104691: gap of unknown length  
104691 109536: contig of 4845 bp in length  
109536 109636: gap of unknown length  
109636 115061: contig of 5425 bp in length  
115061 115161: gap of unknown length  
115161 118242: contig of 3763 bp in length  
118242 119024: gap of unknown length  
119024 123214: contig of 4190 bp in length  
123214 126721: gap of unknown length  
126721 126821: contig of 3407 bp in length  
126821 130079: contig of 3258 bp in length  
130079 130179: gap of unknown length  
130179 134535: contig of 4356 bp in length  
134535 134636: gap of unknown length  
134636 138735: contig of 4000 bp in length  
138735 138636: gap of unknown length

Query Match 68.7%; Score 20.6; DB 2; Length 182196;  
Best Local Similarity 85.2%; Pred. No. 53;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 ctggtttaactggggccagagagagga 27  
||||||| ||||||| |||||||  
Db 68432 CTGCTTTAGTTGGCCACGACGCA 68458

AC078812 194281 bp DNA linear HTG 31-DEC-2000  
Homo sapiens chromosome 3 clone RP11-457K5, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 45 unordered pieces.  
AC078812  
HTG: HTGS\_PHASE1.  
human.  
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 194281)  
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,  
Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Banks, T., Barbara, J.,  
Benton, J., Blinze, K., Blankenburg, K., Bonnin, D., Bouck, J.,  
Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,  
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,  
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,  
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,  
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,  
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,  
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,  
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,  
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,  
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,  
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,  
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,  
Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C.,  
Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,  
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,  
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,  
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,  
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,  
Loulsegh, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,  
Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,  
Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,  
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwenwo, S.,  
Ogih, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,  
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,  
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojebokan, I., Rolfe, M.,  
Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoostari, N.,  
Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,  
Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K.,  
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,  
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,  
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
Watlington, S., Williams, G., Williamson, A., Wlezyk, R., Wooden, S.,  
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.  
and Gibbs, R.  
Direct Submission  
Unpublished  
2 (bases 1 to 194281)  
Worley, K.C.  
Direct Submission

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE

Mon May 20 10:14:51 2002

ACCESSION AP002514  
VERSION AP002514.3 GI:14517595  
KEYWORDS HTG.  
SOURCE Homo sapiens DNA, clone:CTD-3055E19.

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

Source

Location/Qualifiers

1..202833

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="11"

/map="11q"

/clone="CTD-3055E19"

59273 a 39328 c 40151 g 62884 t 1197 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

23; Conservative

0; Mismatches

4; Indels

0; Gaps

0;

Oy

1 ctggtttaaactggccaggagagga 27

||||| ||||| ||||| |||||

Db 119703 CTGCTTTAGTTGGCCCGCAGCAGGA 119729

||||| ||||| ||||| |||||

Search completed: May 18, 2002, 15:00:07

Job time: 8304 sec

\* 138736 142774: contig of 4039 bp in length  
\* 142775 142774: gap of unknown length  
\* 142776 142774: contig of 4330 bp in length  
\* 142777 142774: gap of unknown length  
\* 142778 142774: gap of unknown length  
\* 142779 142774: contig of 3490 bp in length  
\* 142780 142774: gap of unknown length  
\* 142781 142774: contig of 3277 bp in length  
\* 142782 142774: gap of unknown length  
\* 142783 142774: contig of 3185 bp in length  
\* 142784 142774: gap of unknown length  
\* 142785 142774: contig of 3065 bp in length  
\* 142786 142774: gap of unknown length  
\* 142787 142774: contig of 3542 bp in length  
\* 142788 142774: gap of unknown length  
\* 142789 142774: contig of 1608 bp in length  
\* 142790 142774: gap of unknown length  
\* 142791 142774: contig of 2255 bp in length  
\* 142792 142774: gap of unknown length  
\* 142793 142774: contig of 1576 bp in length  
\* 142794 142774: gap of unknown length  
\* 142795 142774: contig of 170102: gap of unknown length  
\* 142796 142774: contig of 2378 bp in length  
\* 142797 142774: gap of unknown length  
\* 142798 142774: contig of 1429 bp in length  
\* 142799 142774: gap of unknown length  
\* 142800 142774: contig of 1433 bp in length  
\* 142801 142774: gap of unknown length  
\* 142802 142774: contig of 2164 bp in length  
\* 142803 142774: gap of unknown length  
\* 142804 142774: contig of 1585 bp in length  
\* 142805 142774: gap of unknown length  
\* 142806 142774: contig of 1022 bp in length  
\* 142807 142774: gap of unknown length  
\* 142808 142774: contig of 1993 bp in length  
\* 142809 142774: gap of unknown length  
\* 142810 142774: contig of 1471 bp in length  
\* 142811 142774: gap of unknown length  
\* 142812 142774: contig of 1608 bp in length  
\* 142813 142774: gap of unknown length  
\* 142814 142774: contig of 1241 bp in length  
\* 142815 142774: gap of unknown length  
\* 142816 142774: contig of 1392 bp in length  
\* 142817 142774: gap of unknown length  
\* 142818 142774: contig of 1696 bp in length  
\* 142819 142774: gap of unknown length  
\* 142820 142774: contig of 1140 bp in length  
\* 142821 142774: gap of unknown length  
\* 142822 142774: contig of 1013 bp in length  
\* 142823 142774: gap of unknown length  
\* 142824 142774: contig of 1214 bp in length.

FEATURES

Source

Location/Qualifiers

1..194281

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="3"

/clone="RP11-457K5"

58344 a 36841 c 37063 g 57557 t 4476 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

23; Conservative

0; Mismatches

4; Indels

0; Gaps

0;

Oy

2 tgggtttaaactggccaggagagga 28

||||| ||||| ||||| |||||

Db 141879 TGGTCTAAAGTGGGGCAGGATAGGAG 141853

||||| ||||| ||||| |||||

Search completed: May 18, 2002, 15:00:07

Job time: 8304 sec

RESULT 15  
AP002514  
LOCUS AP002514 202833 bp DNA linear PRI 21-JUN-2001  
DEFINITION Homo sapiens genomic DNA, chromosome 11q, clone:CTD-3055E19,  
complete sequence.

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 14:58:11 ; Search time 2878.96 Seconds  
(without alignments)  
145.376 Million cell updates/sec

Title: us-09-623-329-22

Perfect score: 20

Sequence: 1 caggttcacgcgcagctcc 20

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hcg.\*

3: gb\_in.\*

4: gb\_on.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pi.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_on.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pi.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htgo\_inv.\*

# SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Match	Length	DB	ID	Description
-----						

1	20	100.0	20	6	AX018463
2	20	100.0	172281	14	EBV
3	20	100.0	184113	14	HS4B958RAJ
4	17.4	87.0	8222	9	AB031010
5	17	85.0	2457	3	DMU56080
6	17	85.0	3175	3	DMU49446
7	17	85.0	82631	3	AE003675
8	17	85.0	145786	3	AE003675
9	17	85.0	157141	3	AC016445
10	16.8	84.0	7756	9	AF065396
11	16.8	84.0	11093	1	AE008272
12	16.8	84.0	13633	1	AE009342
13	16.8	84.0	37641	3	AE005766
14	16.8	84.0	40339	3	AC005767
15	16.8	84.0	58564	2	AC102139
16	16.8	84.0	107025	2	AL139235_3
17	16.8	84.0	133447	2	AC092132
18	16.8	84.0	142667	10	AL135758
19	16.8	84.0	162274	2	AC013819
20	16.8	84.0	162507	9	AC090944
21	16.8	84.0	165134	2	AC103494
22	16.8	84.0	166075	2	AC092836
23	16.8	84.0	175302	2	AC020587
24	16.8	84.0	186271	9	AC024168
25	16.4	82.0	45704	1	AE001827
26	16.4	82.0	70311	9	AC002390
27	16.4	82.0	97825	2	AC024642
28	16.4	82.0	107946	2	AC093945
29	16.4	82.0	114467	9	AP000462
30	16.4	82.0	132513	9	AC004520
31	16.4	82.0	147452	8	AC026815
32	16.4	82.0	154236	2	AC051650
33	16.4	82.0	157399	2	AP003167
34	16.4	82.0	158539	2	OSJN00082
35	16.4	82.0	164815	2	AC090470
36	16.4	82.0	170965	9	AB017654
37	16.4	82.0	171065	2	AC011216
38	16.4	82.0	173907	9	AC022333
39	16.4	82.0	174875	2	AC024682
40	16.4	82.0	175454	9	AF003503
41	16.4	82.0	188740	2	AC024045
42	16.4	82.0	192730	2	AC026676
43	16.4	82.0	194541	2	AC011119
44	16.4	82.0	214623	2	AC069209
45	16.4	82.0	331806	2	AC022365

## ALIGNMENTS

RESULT 1

AX018463

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. .20

/organism="Human herpesvirus 4"

/db\_xref="taxon:10376"

3 a 8 c

BASE COUNT

5 t

AX018463 Sequence 20 bp DNA linear PAT 07-SEP-2000

Sequence 22 from Patent WO945155.

AX018463

AX018463.1 GI:10042614

Human herpesvirus 4

Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

Gammaherpesvirinae; Lymphocryptovirus.

1 (bases 1 to 20)

Middelborg, J.M., Van Den Brule, A.J. and Vervoot, M.B.

Oligonucleotides for the amplification and detection of Epstein

barr virus (ebv) nucleic acid

Patent: WO 945155-A 22 10-SEP-1999;

MIDDELBOURG JAAP MICHIEL (NL); AKZO NOBEL NV (NL); DEN BRULE

ADRIANUS JOHANNES CH (NL); VERVOORT MARCEL BARTOLINA HEND (NL)

Location/Qualifiers

1. .20

/organism="Human herpesvirus 4"

/db\_xref="taxon:10376"

3 a 8 c

BASE COUNT

5 t

ORIGIN	Query Match Best Local Similarity 100.0%; Score 20; DB 6; Length 20; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	The EMBO journal. 2 (8), 1331-1338 (1983) 20331131 10872327 8 (bases 142687 to 159853) Bankier,A.T., Deininger,P.L., Farrell,P.J. and Barrell,B.G. Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8 Epstein-Barr virus Molecular biology & medicine. 1 (1), 21-45 (1983) 85035713 6092825 9 (bases 112620 to 125316) Sequin,C., Farrell,P.J. and Barrell,B.G. DNA sequence and transcription of the BamHI fragment B region of B95-8 Epstein-Barr virus Molecular biology & medicine. 1 (3), 369-392 (1983) 85060424 6094953 10 (bases 45415 to 52824) Jones,M.D., Foster,L., Sheedy,T. and Griffin,B.E. The EB virus genome in Daudi Burkitt's lymphoma cells has a deletion similar to that observed in a non-transforming strain (p3HR-1) of the virus The EMBO journal. 3 (4), 813-821 (1984) 84207939 6327290 11 (bases 87650 to 92703) Biggin,M., Farrell,P.J. and Barrell,B.G. Transcription and DNA sequence of the BamHI L fragment of B95-8 Epstein-Barr virus The EMBO journal. 3 (5), 1083-1090 (1984) 84236104 6203743 12 (bases 7315 to 9312) Yates,J., Warren,N., Reisman,D. and Sugden,B. A cis-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells Proceedings of the National Academy of Sciences of the United States of America. 81 (12), 3806-3810 (1984) 84122045 6328526 13 (bases 76089 to 79808) Gibson,T., Stockwell,P., Ginsburg,M. and Barrell,B. Homology between two EBV early genes and HSV ribonucleotide reductase and 38K genes Nucleic acids research. 12 (12), 5087-5099 (1984) 84247360 6330697 14 (bases 1 to 172281) Baer,R.J., Bankier,A.T., Biggin,M.D., Deininger,P.L., Farrell,P.J., Gibson,T.J., Hattuli,G.F., Hudson,G.S., Satchwell,S.C., Seguir,C., Tufnell,P.S. and Barrell,B.G. DNA sequence and expression of the B95-8 Epstein-Barr virus genome Nature. 310 (5974), 207-211 (1984) 84270667 6087149 15 (bases 1 to 172281) Bodscot,M. and Perricaudet,M. Clustered alternative splice sites in Epstein-Barr virus RNAs Nucleic acids research. 15 (14), 5887 (1987) 87289053 3039467 16 (bases 1 to 172281) Lauk,G., Perricaudet,M. and Farrell,P.J. A spliced Epstein-Barr virus gene expressed in immortalized lymphocytes is created by circularization of the linear viral genome The EMBO journal. 7 (3), 769-774 (1988) 88283646 2840285 17 (bases 1 to 172281) Hattuli,G.F., Barrell,B.G., Quinn,J. and McGeoch,D. Unpublished
Qy	1 cagggttcacgtcagctcc 20 	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
Db	1 CAGGTTTCATCGCTCAGCTCC 20 	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
RESULT 2		JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
EBV		JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
LOCUS		JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
DEFINITION	172281 bp DNA circular VRL 20-SEP-1999	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
ACCESSION	Epstein-Barr virus (EBV) genome, strain B95-8.	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
VERSION	V01555 J02070 K01729 K01730 V01554 X00498 X00499 X00784	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
KEYWORDS	V01555.1 GI:59074 DNA polymerase; EBNA: genome; ribonucleotide reductase; tandem repeat; terminal repeat.	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
SOURCE	Epstein-Barr virus.	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
ORGANISM	Human herpesvirus 4	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
REFERENCE	Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammapherpesvirinae; Lymphocryptovirus.	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
AUTHORS	1 (bases 1 to 172281) Arrand,J.R., Rymo,L., Walsh,J.E., Bjorck,E., Lindahl,T. and Griffin,B.E.	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
TITLE	Molecular cloning of the complete Epstein-Barr virus genome as a set of overlapping restriction endonuclease fragments	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
JOURNAL	Nucleic acids research. 9 (13), 2999-3014 (1981)	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
MEDLINE	82014887	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
PUBMED	6269068	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
REFERENCE	2 (bases 1 to 172281)	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
AUTHORS	Kozak,M.	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
TITLE	Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
JOURNAL	Nucleic acids research. 9 (20), 5233-5262 (1981)	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
MEDLINE	82059504	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
PUBMED	7301588	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
REFERENCE	3 (bases 1 to 172281)	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
AUTHORS	Deininger,P.L., Bankier,A., Farrell,P., Baer,R. and Barrell,B.	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
TITLE	Sequence analysis and in vitro transcription of portions of the Epstein-Barr virus genome	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
JOURNAL	Journal of cellular biochemistry. 19 (3), 267-274 (1982)	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
MEDLINE	83109311	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
PUBMED	6296170	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
REFERENCE	4 (bases 1 to 172281)	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
AUTHORS	Farrell,P.J., Deininger,P.L., Bankier,A. and Barrell,B.	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
TITLE	Homologous upstream sequences near Epstein-Barr virus promoters Proceedings of the National Academy of Sciences of the United States of America. 80 (6), 1565-1569 (1983)	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
JOURNAL	83169725	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
MEDLINE	6300857	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
REFERENCE	5 (bases 45644 to 52450)	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
AUTHORS	Jiang,K.T. and Hayward,S.D.	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
TITLE	Organization of the Epstein-Barr virus DNA molecule. III. Location of the p3HR-1 deletion junction and characterization of the NotI repeat units that form part of the template for an abundant 12-O-tetradecanoylphorbol-13-acetate-induced mRNA transcript Journal of virology. 48 (1), 135-148 (1983) 83294686 6310141	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
JOURNAL	83294686	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
MEDLINE	6310141	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
REFERENCE	6 (bases 159853 to 172281)	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
AUTHORS	Bankier,A.T., Deininger,P.L., Satchwell,S.C., Baer,R., Farrell,P.J. and Barrell,B.G.	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
TITLE	DNA sequence analysis of the EcoRI Dhet fragment of B95-8 Epstein-Barr virus containing the terminal repeat sequences Molecular biology & medicine. 1 (4), 425-445 (1983) 85060428 6094955	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
JOURNAL	85060428	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
MEDLINE	6094955	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
REFERENCE	7 (bases 1 to 172281)	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
AUTHORS	Farrell,P.J., Bankier,A., Sequin,C., Deininger,P. and Barrell,B.G.	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE
TITLE	Latent and lytic cycle promoters of Epstein-Barr virus	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE

REFERENCE 18 (bases 1 to 172281)  
 AUTHORS Farrell, P.J. and Barrell, B.G.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-JUN-1984)  
 REFERENCE 19 (bases 1 to 172281)  
 AUTHORS Farrell, P.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-MAR-1988) Farrell P., Ludwig Institute for Cancer Research, St. Mary's Hospital Medical School, Norfolk Place London W2 1PG

COMMENTS  
 CDS  
 Listed under this feature are all known protein coding regions as well as all the major open reading frames in the sequence. In general the term major is taken as the longest frame in a particular region taking into account the adjacent longest frames and likely transcription signals. Note that on this basis some long overlapping frames have been excluded and on the other hand some small frames have been included which might represent exons or genes because they occur in a logical combination with other features or because of some other experimental data. The reading frames are named according to the Bam HI fragment in which they start. eg BAlF3 is the third leftward frame starting in Bam HI fragment A. BOPF1 is the first rightward frame in Bam HI fragment O. If there is an obvious TATA sequence followed by an in frame Met codon that satisfies the rules of Kozak [12] in that there is a purine at -3 and/or a G at +4 then the reading frame is numbered from the A of the ATG to the base preceding the termination codon. If there is no obvious initiation codon or there is a substantial reading frame in phase before the ATG then the reading frame is numbered from the first base of the first codon.

SITES OF POLYA signals  
 This feature lists all occurrences of the sequence AATAAA which is found normally approximately 20 bases upstream of the mRNA processing/polyA addition site. The rarely used homolog AITAAA is only listed when it is found in a position close to the end of a major reading frame.

SITES OF DONOR and ACCEPT sequences  
 This is not a comprehensive listing of all such sequences and only the positions of a few have been noted because they occur in potentially interesting positions. The number quoted in the table is the position of the terminal base in the intron in each case. Restriction enzyme SITES.  
 Only the positions of the sites Bam HI (BAM) are listed.

RPT  
 This feature is used to define repetitive sequences.

SITE DEL  
 This feature defines deletions in B95-8 with respect to other strains such as RAJI and also to deletions in other strains such as P3HR1 and DAUDI with respect to B95-8.

SITE HPN  
 Denotes sequences with twofold symmetry ie could form hairpin loops. This is not a comprehensive list - only a few occurrences noted.

ORIGIN  
 Denotes the region that encompasses an origin of replication (ori p). [13].

NUMBERING  
 The DNA sequence of B95-8 EBV has been revised [19]. The original (Baer et al, 1984) base 359 has been deleted so the new sequence around that position reads TCAGTCTTT. To avoid renumbering the entire sequence, position 1 has been moved 1 base to the left of the EcoRI site separating EcoRI Dhet from EcoRI I (ie the first A of AGAATTC).

FEATURES  
 source  
 1. 172281  
 /organism="Human herpesvirus 4"  
 /strain="B95-8"  
 /db\_xref="taxon:10376"

Query Match 100.0%; Score 20; DB 14; Length 172281;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggttcacgtcgcagctcc 20  
 |||||  
 Db 165509 CAGGTTTCAGCTCAGCTCC 165528

RESULT 3  
 HS4B958RAJ 184113 bp DNA linear VRL 12-APR-1996  
 LOCUS Epstein-Barr virus, artifactual joining of B95-8 complete genome  
 DEFINITION and the sequences from Raji of the large deletion found in B95-8.  
 ACCESSION M80517 M75989  
 VERSION M80517.1 GI:330330  
 KEYWORDS Human herpesvirus 4 DNA.  
 SOURCE Human herpesvirus 4  
 ORGANISM Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae; Lymphocryptovirus.

REFERENCE 1 (sites)  
 AUTHORS Baer, R.J., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G.F., Hudson, G.S., Satchwell, S.C., Seguin, C., Tuffnell, P.S. and Barrell, B.G.  
 TITLE DNA sequence and expression of the B95-8 Epstein-Barr virus genome  
 JOURNAL Nature 310 (5974), 207-211 (1984)  
 MEDLINE 84270667

REFERENCE 2 (sites)  
 AUTHORS Parker, B.D., Bankier, A., Satchwell, S., Barrell, B. and Farrell, P.J.  
 TITLE Sequence and transcription of Raji Epstein-Barr virus DNA spanning the B95-8 deletion region  
 JOURNAL Virology 179 (1), 339-346 (1990)  
 MEDLINE 91021036

REFERENCE 3 (sites)  
 AUTHORS Sample, J., Brooks, L., Sample, C., Young, L., Rowe, M., Gregory, C., Rickinson, A. and Kieff, E.  
 TITLE Restricted Epstein-Barr virus protein expression in Burkitt lymphoma is due to a different Epstein-Barr nuclear antigen 1 transcriptional initiation site  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 88 (14), 6343-6347 (1991)  
 MEDLINE 91296817

REFERENCE 4 (bases 1 to 184113)  
 AUTHORS Jenson, H.B.  
 TITLE GenBank Curator Program  
 JOURNAL Unpublished (1992)  
 COMMENT The B95-8 genome (V01555) has a large deletion in the right side of the genome which has been sequenced in Raji (M35547). These sequences have been joined to form an extended and more complete, although artifactual, EBV sequence.

FEATURES  
 source  
 1. 184113  
 /organism="Human herpesvirus 4"  
 /db\_xref="taxon:10376"  
 misc\_feature 1. 152008  
 /note="B95-8 sequences (corresponds to 1-152,008 of V01555)"  
 misc\_feature 152009..152012  
 /note="Overlap of B95-8 and Raji sequences at B95-8 deletion point (corresponds to 152,009-152,012 in V01555, and 1-4 in M35547)"  
 misc\_feature 153013..163839  
 /note="Raji sequences (corresponds to 5-11,831 of M35547)"  
 misc\_feature 163840..163843  
 /note="Overlap of B95-8 and Raji sequences at B95-8 deletion point (corresponds to 152,009-152,012 of V01555, and 11,832-11,835 of M35547)"  
 misc\_feature 163844..184113  
 /note="B95-8 sequences (corresponds to 152,013-172,282 of V01555)"

BASE COUNT 36002 a 55824 c 54622 g 37665 t  
 ORIGIN

Query Match 100.0%; Score 20; DB 14; Length 184113;  
 Best Local Similarity 100.0%; Pred. No. 16;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggttcacgcctcacc 20  
|||||  
Db 177340 CAGGTTTCATCGCTCAGCTCC 177359

RESULT 4  
LOCUS AB031010 8222 bp DNA linear PRI 06-SEP-2000  
DEFINITION Homo sapiens DNA, MHC class I region, 57.1 ancestral haplotype.  
ACCESSION AB031010  
VERSION AB031010.1 GI:7008035  
SOURCE Homo sapiens (haplotype:MHC 57.1 Ancestral Haplotype HLA-A1 HLA-B57 HLA-DR7) Adult B Lymphoblastoid DNA.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Gaudieri,S., Leelayuwat,C., Townend,D.C., Kulski,J.K. and Dawkins,R.L.  
TITLE Genomic characterization of the region between HLA-B and TNF: implications for the evolution of multicopy gene families  
J. Mol. Evol. 44 Suppl 1, S147-S154 (1997)

JOURNAL 97224639  
MEDLINE 2 (bases 1 to 8222)  
REFERENCE Dawkins,R.L. and Gaudieri,S.  
AUTHORS Direct Submission  
TITLE Submitted (12-AUG-1999) to the DBJ/EMBL/GenBank databases. Roger L Dawkins, University of Western Australia, Centre for Molecular Immunology and Instrumentation; Stirling Hwy, Nedlands, WA 6009, Australia (E-mail:sgaudier@lab.nig.ac.jp, Tel:61-8-93464414, Fax:61-8-93464578)  
JOURNAL

COMMENT Sequence incomplete and contains sequence with only single strand coverage.

Sequence contains some ambiguities.

FEATURES Location/Qualifiers

Source 1..8222  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_type="B Lymphoblastoid"  
/chromosome="6"  
/dev\_stage="Adult"  
/haplotype="MHC 57.1 Ancestral Haplotype HLA-A1 HLA-B57 HLA-DR7"  
/map="6p21.3"  
repeat\_unit 441..1118  
/rpt\_family="Tigger3b"  
repeat\_unit 1119..2176  
/rpt\_family="L1pA3"  
repeat\_unit 2248..2752  
/rpt\_family="Tigger3b"  
repeat\_unit 2862..3165  
/rpt\_family="AluJb"  
repeat\_unit 3367..3628  
/rpt\_family="AluJb"  
repeat\_unit complement(4238..4928)  
/rpt\_family="MER21b"  
repeat\_unit 4932..6197  
/rpt\_family="L1M1"  
repeat\_region 6198..6229  
/note="(TAAAA)n"  
repeat\_unit 6230..6366  
/rpt\_family="L1M1"  
repeat\_region 6958..7005  
/note="(CA)n"  
repeat\_region 7057..7092  
/note="(TCC)n"  
repeat\_unit complement(7097..7132)  
/rpt\_family="THE1b"  
BASE COUNT 2657 a 1705 c 1665 g 2163 t 32 others  
ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 8222;  
Best Local Similarity 94.7%; Pred. No. 3.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 caggttcacgcctcacc 19  
|||||  
Db 6611 CAGGTTTCATCGCTCAGCTC 6629

RESULT 5  
LOCUS DMU56080 2457 bp mRNA linear INV 30-MAY-1996  
DEFINITION Drosophila melanogaster p21-activated protein kinase 1 Dpkl  
(Dpkl) mRNA, complete cds.

ACCESSION U56080  
VERSION U56080.1 GI:1336003

KEYWORDS fruit fly.

SOURCE Drosophila melanogaster  
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 2457)  
AUTHORS Chernoff,J.

TITLE Direct Submission  
JOURNAL Submitted (22-APR-1996) Jonathan Chernoff, Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111, USA

FEATURES Location/Qualifiers

Source 1..2457  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone\_lib="Drosophila 8 hr cDNA library (Brown and Katatos)"  
gene 1..2457  
/gene="Dpkl"  
CDS 134..2248  
/gene="Dpkl"  
/codon\_start=1  
/product="p21-activated protein kinase 1 Dpkl"  
/protein\_id="AAB01209.1"  
/db\_xref="GI:1336004"

/translation="MSSEEDKPPAPVRLTSNRGNERSGGVGVGGGGLGGGGMGV  
PPDMRPLKPEPDDSDRRKKTISKIKGSKPSTDSKPNISYPTNFTVHGVDFATVG  
EFTCMPEAWARLLMNSNISKQKQNFQAVLDLVKWFONTTKQRPSSKYMTNATTHS  
GSSLSRVSSSSPSTPTDSELHGSSGNLIGVSGMTLGPANNAVAGQILCNHY  
QQQQHLLQQOQPLHQNHNHMGISQSHSYNEVHTVSSSTQSHSSANEDMLGPQ  
HPQQPPPPPVASRPRTKSYTRPDELQPAITPMVAPATTTPLONHRTFGGIS  
APATSPMHNNATTTLDKNNANIYTPETVAQVAGSSGSSVAGMOIAVPAAVAP  
AATPNTAANAANKKKMSDEIELEKRTIVSDPNRKYTKMEKIGOGASCTVSAIES  
STGMEVALKQNLISQPKKEILINELLVMRENKHPNVNYLDSYLVSEELWVMVEYLP  
GGSLTDVVTETCMBDEGOIAGVCREVLOALEFLHANQVLIHRDIKSDNILLGLDGSVKLT  
DFGFC AOLSPESQSKTTMGVPMYMAPEVTVTRKOYGPVKDLMSLIGIMAIEMVEGPPY  
LNENPLKALYLIATNGKPEIKKKLSSAFQDFLDQCLEVEVDHRSALDLKLHPFLK  
LARPLASLTPLMAAKKATGN"

BASE COUNT 694 a 688 c 630 g 442 t 3 others  
ORIGIN

Query Match 85.0%; Score 17; DB 3; Length 2457;  
Best Local Similarity 100.0%; Pred. No. 5.9e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggttcacgcctcacc 17  
|||||  
Db 612 CAGGTTTCATCGCTCAGC 628

RESULT 6  
LOCUS DMU49446  
DEFINITION Drosophila melanogaster serine/threonine kinase PAK homolog DPAK

```

mRNA, complete cds.
ACCESSION U49446
VERSION U49446.1 GI:1335889
KEYWORDS
SOURCE fruit fly strain-Canton S.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 3175)
AUTHORS Harden,N., Lee,J., Loh,H.Y., Ong,Y.M., Tan,I., Leung,T., Manser,E.
and Lim,L.
TITLE A Drosophila homolog of the Rac- and Cdc42-activated
serine/threonine kinase PAK is a potential focal adhesion and focal
complex protein that colocalizes with dynamic actin structures
Mol. Cell. Biol. 16 (5), 1896-1908 (1996)
JOURNAL
MEDLINE
REFERENCE 2 (bases 1 to 3175)
AUTHORS Harden.N.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-1996) Nicholas Harden, IMCB, National University
of Singapore, 10 Kent Ridge Crescent, Singapore 119260, Singapore
FEATURES
Source
1..3175
/organism="Drosophila melanogaster"
/strain="Canton S"
/db_xref="taxon:7227"
/chromosome="83E"
1..3175
/gene="DPAK"
269..2383
/gene="DPAK"
/codon_start=1
/product="serine/threonine kinase PAK homolog DPAK"
/protein_id="AAC47094.1"
/db_xref="GI:1335890"
/translation="MSSEEDKPPAPPVRLTSNRGNGRSGGVGGVGGGLGGGMDV
PDMRLPEPDSDRKKTKLTKSGKPSHTDSKPNISYPTNEHTVHVGFDVTV
ETGPEAWARLLMSNTLSKQKKNPOAVLDVLFKDFONTKORPSKYMTNATTHS
GSLSRVSSPSSPTTSELHSGSGNLGVGLSGMTLGNANNAVAGDILNGHY
QQQQHLLQQQPLHONHNMHMGISQSHSNFVGHTVSSSTQSHSSANDMLGPG
HPQQPPPPVSPASPERTKSIYTRFIEDLPALIPMPAPATTPATPQNHRTGGIS
APATSPMHNNATTTLDKNNKNANLYTPTPTVAQVSGPSSQVAGNQIAPQAAVAP
LATPNRAAKNKKMSDEEILEKLTIVSGDPPNRKVTMEKIGOGASGTVTYTAIES
STGMEVAIKOMLSQPKKELIINEITVMRENKHPNVNYLDSYLSBELVWMEVLP
GSLTDVVTETCDGQIACVCEVLQALEFLHANQVLRDILKSDNILLGLDGSVKLT
DFGCAQISPEQSKRTTWTGTPYMAPEVTRKQYGPVKVDLSLGLTMAIEMVEGPPY
LNENPLASLTLIATNGKPEIKERKDLSSAFQDFLQOCLEVEVDRRASALDLKHFFLK
LARPLASLTPLMAAKEATKGN"
BASE COUNT 959 a 811 c 740 g 665 t
ORIGIN
Query Match 85.0%; Score 17; DB 3; Length 3175;
Best Local Similarity 100.0%; Pred. No. 5.8e-02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggttcacgcctcagc 17
|||||
Db 747 CAGGTTTCATCGCTCAGC 763

RESULT 7
AC020316/c
LOCUS AC020316 82631 bp DNA linear HTG 03-JAN-2000
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC020316
VERSION AC020316.1 GI:6664581
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 82631)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10212745 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Source
1..82631
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 23613 a 17309 c 17873 g 23836 t
ORIGIN
Query Match 85.0%; Score 17; DB 2; Length 82631;
Best Local Similarity 100.0%; Pred. No. 5.4e-02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 caggttcacgcctcagc 17
|||||
Db 60383 CAGGTTTCATCGCTCAGC 60367

RESULT 8
AC003675
LOCUS AC003675 145786 bp DNA linear INV 04-OCT-2000
DEFINITION Drosophila melanogaster genomic scaffold 14200001386040 section 1
of 5, complete sequence.
ACCESSION AE003675 AE002699
VERSION AE003675.2 GI:10727124
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 145786)
AUTHORS Adams,M.D., Celnikier,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.M.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,J.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,I., Beasley,E.M., Beeson,K.Y.,
Bencos,P.V., Bereman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,C., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahike,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
Fleischmann,W., Fosler,C., Gabriellian,A.E., Garq,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennisson,J., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,

```

Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,  
 Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,  
 Stapleton,M., Strong,R., Sun,E., Swirskas,R., Tector,C., Turner,R.,  
 Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,  
 Weinstein,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,  
 Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,  
 Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,  
 Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,  
 Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.  
 The genome sequence of *Drosophila melanogaster*  
 Science 287 (5461), 2185-2195 (2000)  
 20196006  
 2 (bases 1 to 145786)  
 Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.  
 Direct Submission  
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD, USA  
 On Oct 9, 2000 this sequence version replaced gi:7298913.  
 COMMENT Location/Qualifiers  
 FEATURES  
 source  
 1..145786  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /chromosomes="3r"  
 join(<946..1146,1425..1701,1783..1964,2303..2443)  
 /gene="CG15189"  
 /product="CT35104"  
 /db\_xref="FLYBASE:FBan0015189"  
 /db\_xref="FLYBASE:FBgn0037429"  
 /evidence=not\_experimental  
 <946..>2443  
 /gene="CG15189"  
 /map="83E4-83E4"  
 /db\_xref="FLYBASE:FBan0015189"  
 /db\_xref="FLYBASE:FBgn0037429"  
 /evidence=not\_experimental  
 join(946..1146,1425..1701,1783..1964,2303..2443)  
 /gene="CG15189"  
 /note="CG15189 gene product"  
 /codon\_start=1  
 /db\_xref="FLYBASE:FBan0015189"  
 /db\_xref="FLYBASE:FBgn0037429"  
 /evidence=not\_experimental  
 /protein\_id="AAF54139.1"  
 /db\_xref="GI:7298934"  
 /translation="MAKLLLVIGVVAALVAQAAGSGTEKMRILAEQNKACSGQD  
 SMACIKERAMREVQVMSKDSQVSNLEVRNCKETPTINERASSAGGFLDAIENYI  
 RGDVSMIDLPLADAKVTVSARNLVNLSNIQI.NGDDGDECTDVEARGKGNIEKKG  
 KKHRLKAMPILVILILKAITVIPAIGILKIKAFNALALGFFSFIYVGLAIFQLC  
 KTIADHHTAHTAHGPMWDRFTSSVPAPVVEQPKLGQALAYQAYA"  
 join(<5692..6094,6161..>6600)  
 /gene="CG15188"  
 /product="CT35103"  
 /db\_xref="FLYBASE:FBan0015188"  
 /db\_xref="FLYBASE:FBgn0037430"  
 /evidence=not\_experimental  
 <5692..>6600  
 /gene="CG15188"  
 /map="83E4-83E4"  
 /db\_xref="FLYBASE:FBan0015188"  
 /db\_xref="FLYBASE:FBgn0037430"  
 /evidence=not\_experimental  
 join(5692..6094,6161..6600)  
 /gene="CG15188"  
 /note="CG15188 gene product"  
 /codon\_start=1  
 /db\_xref="FLYBASE:FBan0015188"  
 /db\_xref="FLYBASE:FBgn0037430"  
 /evidence=not\_experimental  
 /protein\_id="AAF54138.1"  
 /db\_xref="GI:7298933"  
 /translation="WAFSTSLAFSCALLLVASTVSQAATENAVTPRIHSSDELIS  
 TVDCKCFHANMCHLKEKLVLTVDIVANVEEVSGRALGDDKVIDKIVDRILNTN  
 EMRLQLPQTFAGSVVYTSRDRGDELDPKDEGRKKNKLFPLPLLLMKFLKVI

MPILLALIGLKATKALILSKIAIKLVLCGLIYNILOKLGCMKMWPMPPADVPASEYC  
 VPSITASSYDPSSEWPMGGPYARWDSNLAISYTHPSSSSSSSSSGSSSGSSSSS  
 SSSS"  
 join(<7505..7639,7840..>8340)  
 /gene="CG17917"  
 /product="CT39914"  
 /db\_xref="FLYBASE:FBan0017917"  
 /db\_xref="FLYBASE:FBgn0037431"  
 <7505..>8340  
 /gene="CG17917"  
 /map="83E4-83E5"  
 /db\_xref="FLYBASE:FBan0017917"  
 /db\_xref="FLYBASE:FBgn0037431"  
 join(7505..7639,7840..8340)  
 /gene="CG17917"  
 /note="CG17917 gene product"  
 /codon\_start=1  
 /db\_xref="FLYBASE:FBan0017917"  
 /db\_xref="FLYBASE:FBgn0037431"  
 /protein\_id="AAG22206.1"  
 /db\_xref="GI:10727125"  
 /translation="MRFAPSLSLKSTRGVHQSDETSVKIMRSLDVIDVTHIGQREFL  
 NMYTHGLAHGCGKVLPEMQVRDEPSVKMSPAPENYALLMVDVDPVNAITPTTHREFL  
 HWYVNLTPGNLLALGDVRYGMYGATPLKGTGTHREFVLLYKORDYTKDFPKLPKHSV  
 KGRSGFTKRFKKYRFCHPVAGNFFTSQWSPDVPSLIKAISHNARQVAHF"  
 join(8553..8654,8708..9282)  
 /gene="CG10298"  
 /product="CT28915"  
 /db\_xref="FLYBASE:FBan0010298"  
 /db\_xref="FLYBASE:FBgn0037432"  
 <8553..>9282  
 /gene="CG10298"  
 /map="83E4-83E5"  
 /db\_xref="FLYBASE:FBan0010298"  
 /db\_xref="FLYBASE:FBgn0037432"  
 join(8574..8654,8708..9190)  
 /gene="CG10298"  
 /note="CG10298 gene product"  
 /codon\_start=1  
 /db\_xref="FLYBASE:FBan0010298"  
 /db\_xref="FLYBASE:FBgn0037432"  
 /protein\_id="AAF54137.1"  
 /db\_xref="GI:7298932"  
 /translation="MSDSTVCFSKHKIVPDILKTCPTALLTVTYGGQVVDVGGELTP  
 TOVSQPKVKWDADNAFYETILLTDPDAPSRKPKFREMHHMLVYNIIPGNQVENGWVL  
 DYVAGPQGTGLHRYVFLVFKQPKILCTCNEPKIPKTSQDKRANFSTSKPMKYKLG  
 DPYAGNFFQAQWDDYVPLKYQLSGCK"  
 join(9550..9753,10035..10600)  
 /gene="CG17919"  
 /product="CT39916"  
 /db\_xref="FLYBASE:FBan0017919"  
 /db\_xref="FLYBASE:FBgn0037433"  
 <9550..>10600  
 /gene="CG17919"  
 /map="83E4-83E5"  
 /db\_xref="FLYBASE:FBan0017919"  
 /db\_xref="FLYBASE:FBgn0037433"  
 join(9625..9753,10035..10514)  
 /gene="CG17919"  
 /note="CG17919 gene product"  
 /codon\_start=1  
 /db\_xref="FLYBASE:FBan0017919"  
 /db\_xref="FLYBASE:FBgn0037433"  
 /protein\_id="AAF54136.1"  
 /db\_xref="GI:7298931"  
 /translation="MLRVLLPLVGLCLAVAGSVEEVRSHOVVDVPEPPNOLLKV  
 TYNNLVKDGVELTPTQVKQPVVWDAQGEFYTLTMDPDAPSRAPKFRREFKHW  
 ILANIAGNDLASEGPIAEYIGSGPQGLHRYVFLYKQSGKLEFDEERVSKRSRKD  
 RPKFSAAKFAINHELGNFIAGTYQAOYDDVVPKHLKOLSEN"  
 complement(join(10720..11139,11274..11373))  
 /gene="CG1249"  
 /product="CT2462"  
 /db\_xref="FLYBASE:FBan0001249"



```

/db_xref="FLYBASE:FBgn0037434"
complement(<10720..>.11373)
/genes="CG1249"
/map="83E4-83E5"
/db_xref="FLYBASE:FBan0001249"
/db_xref="FLYBASE:FBgn0037434"
complement(join(10783..11139,11274..11276))
/genes="CG1249"
/note="CG1249 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0001249"
/db_xref="FLYBASE:FBgn0037434"
/protein_id="AAF54135.1"

Query Match      85.0%; Score 17; DB 3; Length 145786;
Best Local Similarity 100.0%; Pred. No. 5.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggttcacgcgcagc 17
|||||
Db 22251 CAGGTCATCGCTCAGC 22267

RESULT 9
AC016445/c
AC016445 LOCUS
DEFINITION Drosophila melanogaster, chromosome 3R, region 83D-83F, BAC clone
BACR19H03, complete sequence.
AC016445
ACCESSION AC016445.4 GI:15624854
VERSION HTG.
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 157141)
AUTHORS Celnikier,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Bazonn,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Ferrera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ikebaw,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunco,J.,
Pacleb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 3R, region 83D-83F
Unpublished
2 (bases 1 to 157141)
AUTHORS Celnikier,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champs,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
Direct Submission
Submitted (30-NOV-1991) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Sep 15, 2001 this sequence version replaced gi:6633929.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgpf@fruitfly.berkeley.edu.
Location/Qualifiers
1..157141
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="83D-83F"
/clone="BACR19H03 (D1184)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial ECORI in
pBACE3.6)"
BASE COUNT 43435 a 32580 c 34482 g 46644 t
ORIGIN

FEATURES
source
1..157141
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="83D-83F"
/clone="BACR19H03 (D1184)"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial ECORI in
pBACE3.6)"
BASE COUNT 43435 a 32580 c 34482 g 46644 t
ORIGIN

Query Match      85.0%; Score 17; DB 3; Length 157141;
Best Local Similarity 100.0%; Pred. No. 5.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caggttcacgcgcagc 17
|||||
Db 33548 CAGGTCATCGCTCAGC 33532

RESULT 10
AF065396 LOCUS
DEFINITION Homo sapiens retinoic X receptor B gene, complete cds.
ACCESSION AF065396
VERSION AF065396.1 GI:3172497
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 7756)
AUTHORS Numasawa,T., Koga,H., Ueyama,K., Maeda,S., Sakou,T., Harata,S.,
Leppert,M. and Inoue,I.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-1998) Molecular Genetics, IMCR, Gunma University,
Showa-machi 3-39-15, Maebashi, Gunma 371, Japan
FEATURES
source
1..7756
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="6p21"
join(676..1082,1908..2155,2857..3013,3380..3559,
4720..4892,5290..5419,5624..5756,5872..5963,6276..6381,
6497..6880)
/product="retinoic X receptor B"
/exon 676..1082
/number=1
join(848..1082,1908..2155,2857..3013,3380..3559,
4720..4892,5290..5419,5624..5756,5872..5963,6276..6381,
6497..6644)
/codon_start=1
/product="retinoic X receptor B"
/protein_id="AAC18599.1"
/db_xref="GI:3172498"

TRANSLATION="MSWAARPPPLPORHAAGCGPGVGVKEMHCGVASWRNRPPWLD
PRAAAAAGVGEQQTPEPEAGRDGMDGRSDSSPNPLPGVPPPPSPG
PLPSTAPSLAGSGAPPMPPPPLGSPFPVYISSMGSGLPPPPGSGPSSP
QINSTVSLPGSGGPPEDVPPVLVGRGLHCPPPGPGAGKRLCAICDRSGSKHYG
VYSCGCKGFKKRTIRKDLTYSCRNKDKTVDKRNRCOYCRQKCLATGMKREAVQ
EERQKDKDGDGEGAGPEEMPVDRILEAEAVQKSDQGVGPGGTGGSGSSPND
PVYNIQAADKQLFTLVWAKRIPHFSSLPDQVILLRAGNELLIAFSHRSIDVR
DGLLATGLVHRHNSAHGVAIFDRVITELVSKMRDMRKTELGCLRAILLFPD
AGKLSNPSVEVLREKYASLETYCKYQKYPEQOGRFAKLLLRPALRSLGLKCLEHLF
FFKLIGDTIDTFLMELMEAPHQLA"
1908..2155

```



```

CDS
/gene="AGR_L1475"
/complement(5191..6030)
/gene="AGR_L1475"
/feature="probable ATP-binding component of ABC transporter
PA4505 [imported] - Pseudomonas aeruginosa (strain PAO1)"
/codon_start=1
/transl_table=11
/product="AGR_L1475p"
/protein_id="AAK89315.1"
/db_xref="GI:15159153"
/translation="MSLLKIKLTVKFAATCAVTAAGIDVSDKHVEYLAIVGESGS
GKSVSLAVMGLLPDTATITADEMTFDGKNNLMSQERRVIGREITMIFQEPVASL
NFSFTGFTTEVRLNLCMGAAARARLLELFQAVGPEPEPTKLKAKPHQMSGGQCQ
RVNIAIAISKPRLLIATPTALDVTIQKILDLMLKLEEYGMALILITDHMGVVA
ETADRVVQYKGRKMEADVLSLFEAPQHPTKALISALPENATGDRLPTVSDFFGCE
GVP"
/complement(6034..6948)
/gene="AGR_L1477"
/complement(6034..6948)
/gene="AGR_L1477"
/feature="probable permease of ABC transporter PA4504
[imported] - Pseudomonas aeruginosa (strain PAO1)"
/codon_start=1
/transl_table=11
/product="AGR_L1477p"
/protein_id="AAK89316.1"
/db_xref="GI:15159154"
/translation="MTDVIITPGVRLSDAAIRRRMIADFWYFRQNRGAVIGLAVFIL
LVLIYFAAVIAPHDPTQYRDALLVPVQMDGGAFLGTDGVRGDMLSRLIYGAQ
YSLFGVNTAIVAGVIGVIGLVAGFFGWDVTVINRVMDVILAPFSLLLALVLAIVL
GGLTNMTAIVAGVIGVIGLVAGFFGWDVTVINRVMDVILAPFSLLLALVLAIVL
LAPLIVQATLSFSSAVLDAAALGFLGNGQAPPAENGWTLAFAREFILRAVWVTVLPG
LTLIIIVLAINMGDLGRDLDPKLRSS"
/complement(6945..7952)
/gene="AGR_L1478"
/complement(6945..7952)
/gene="AGR_L1478"
/feature="DIPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN DPPB"
/codon_start=1
/transl_table=11
/product="AGR_L1478p"
/protein_id="AAK89317.1"
/db_xref="GI:15159155"
/translation="MIRFIFGKLLVLPPTFLGITIVAFAPVRVLPDGPVLLMAGERGI
SPERHAQALQGLFQAPIMQVYHLHFLGRLLQDGLNSLVTGKPKVLTEFLALPATVEL
GLVATIAIVGVGVVIAAIRRGSNFQDQISMTALVGSMPFIWGLLLIIMFSGIL
QWTPVSGRISLMTFFSVTGFMILDSISQKGFASALSHLLPLSPVYVLTATPLAVIA
ROTRSAMLEVLGDEYVTRAKGMSASRVVGIHALRNAMIPVITITGLQIGVLMAGAI
LTETIFSPGIGKWMIDISIRRDYPPVQSGLLLIAGLVVWVNVLLVDLTYGLINPRIH
K"
/complement(8015..9619)
/gene="AGR_L1481"
/complement(8015..9619)
/gene="AGR_L1481"
/feature="dipeptide transport protein"
/codon_start=1
/transl_table=11
/product="AGR_L1481p"
/protein_id="AAK89318.1"
/db_xref="GI:15159156"
/translation="MKFLTRSLIIASALAMASAMPAMAKTFVYCSEASPSGPDSPYT
AGGTFDASHPVNRUTEFKKGTTEVEPGLAEKWDVSSDGLTYTHLRKGVKFEHNDK
FTPSRDPNADVDIFSNRQDGTKNPNQYIAGITYEYNSMPSLKIETVYKVDYTV
KFVLTREPADFLANIPFASIVSKEYADTLEKAGTEKEDFNPLIGTGPKEFVAYOKD
AVTRYOKNADWDGAPKIDDLIFAITPDAARLQKLAGECHELMYPAPADLSTIRAD
KNLIDROPGLNVAFYNTVAFQKPEVRKALNMAMKQALIDAVFCQAGVAKNP
IPTTWISYNDISIKDDAYDPEAKKALEAGYKDLTKIWMAPVQRPYMPNARTAEII
QSDFAKVGVAEIVFEMWETLKRSTEVNRDGAIVLGTGNDGDPDMFNGWFLASCAAT
GEGGANRQCNKEFSSELLSKAQTTDVAERTKLYEQAVIFKEQAPWATLHASTQFV
PMGAKVSGFTWSPGLDFTFENVDAIE"
/complement(9737..10978)
/gene="AGR_L1483"
/complement(9737..10978)

```

```

Query Match      84.0%; Score 16.8; DB 1; Length 11093;
Best Local Similarity 90.0%; Pred. No. 7.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 caggttcacgcgcagctcc 20
      |||| |||| |||| ||||
Db 6022 CAGGCTCATTGCTCAGCTCC 6041

RESULT 12
AE009342/c
LOCUS      13633 bp      DNA      linear      BCT 20-DEC-2001
DEFINITION Agrobacterium tumefaciens str. C58 linear chromosome, section 112
of 187 of the complete sequence.
ACCESSION  AE009342 AE008689
VERSION     AE009342.1 GI:17742556
KEYWORDS
SOURCE      Agrobacterium tumefaciens str. C58 (U. Washington).
ORGANISM    Agrobacterium tumefaciens str. C58 (U. Washington)
REFERENCE   1 (bases 1 to 13633)
AUTHORS     Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
Zhou,Y., Bovee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
Gillet,W., Grant,C., Guenther,D., Kutyavin,T., Levy,R., Li,M.,
McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
and Nester,E.W.
The Genome of the Natural Genetic Engineer Agrobacterium
tumefaciens C58
Science 294 (5550), 2317-2323 (2001)
11743193
REFERENCE   2 (bases 1 to 13633)
AUTHORS     Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
Zhou,Y., Bovee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
Gillet,W., Grant,C., Guenther,D., Kutyavin,T., Levy,R., Li,M.,
McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
and Nester,E.W.
Direct Submission
Submitted (27-SEP-2001) Department of Microbiology, University of
Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA
98195-7242, USA
FEATURES
source      1..13633
             Location/Qualifiers
             /organism="Agrobacterium tumefaciens str. C58 (U.
             Washington)"
             /strain="C58"
             /db_xref="taxon:180835"
             complement(81..1622)
             /gene="Atu4105"

```



IPPTWSYNDISKDAVDPEAAKALEAGVLDLTKIWMVPORBPYNARRTAELI  
QSDPAKVQKAEIVSFEMGELKKSSTEVRDGAIVGLWTGQVDFPMFVLLSCAAT  
GEGANRAQVKNKFSLELSKAKQTDTVAERTKLYEQAFIFKQAPMATLASHQFV  
PMSAKVSGFTMSPLGDTFFENVDIAE\*  
9216..10223  
/gene="dppb"

Query Match 84.0%; Score 16.8; DB 1; Length 13633;  
Best Local Similarity 90.0%; Pred. No. 7.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 caggttcacgtcagctcc 20  
||||| ||||| ||||| |||||  
Db 11146 CAGGCTCATGCTCAGCTCC 11127

RESULT 13  
AC005766/c  
LOCUS  
DEFINITION  
Leishmania major chromosome 3 clone L4625 strain Friedlin, complete  
sequence.  
AC005766  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Leishmania major  
Leishmania major  
Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;  
Leishmania.

REFERENCE  
AUTHORS  
Myler, P.J., Sisk, E., Hixson, G., Kiser, P., Rickel, E., Hassebrock, M.,  
Cawthra, J., Marsolini, F., Sunkin, S. and Stuart, K.D.

TITLE  
JOURNAL  
Submitted (03-09-1998) Seattle Biomedical Research Institution, 4  
Nickerson Street, Seattle, WA 98109-1651, USA

REFERENCE  
AUTHORS  
Myler, P.J., Sisk, E., Hixson, G., Kiser, P., Rickel, E., Hassebrock, M.,  
Cawthra, J., Marsolini, F., Sunkin, S. and Stuart, K.D.

TITLE  
JOURNAL  
Submitted (25-SEP-1999) Seattle Biomedical Research Institution, 4  
Nickerson Street, Seattle, WA 98109-1651, USA

REFERENCE  
AUTHORS  
Myler, P.J., Sisk, E., Hixson, G., Kiser, P., Rickel, E., Hassebrock, M.,  
Cawthra, J., Marsolini, F., Sunkin, S. and Stuart, K.D.

TITLE  
JOURNAL  
Submitted (13-NOV-1999) Seattle Biomedical Research Institution, 4  
Nickerson Street, Seattle, WA 98109-1651, USA

REFERENCE  
AUTHORS  
Myler, P.J.  
Direct Submission  
Submitted (27-JAN-2000) Seattle Biomedical Research Institute, 4  
Nickerson Street, Seattle, WA 98109-1651, USA

COMMENT  
FEATURES  
On Nov 13, 1999 this sequence version replaced gi:5923658.  
Location/Qualifiers

1. 37641  
/organism="Leishmania major"  
/strain="Friedlin"  
/db\_xref="taxon:5664"  
/chromosome="3"  
/clone="L4625"

<1. 1300  
/gene="2AEPAT"  
/gene="2AEPAT"  
/gene="2AEPAT"

/note="BLASTP and COG similarity to Serine-pyruvate  
aminotransferases from several organisms"  
/codon\_start=1  
/product="2-aminoethylphosphonate:pyruvate  
aminotransferase"

/protein\_id="AAF28367.1"  
/db\_xref="GI:6778463"  
/translation="LSFNFRHVESPACPPASFPVADLSITSKKNTHKLLSFFIFPL  
FLESLSVLPQKPLFTPGPLTSDTKREMLTDYASRDINRQAVKFIREDLSVA  
GLDPQKAWLIQQSGSGMGIEAISTIFPTGKFLINSGKYSKQACVWKRLQFPM

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

VLMKGEELKMSALESIRANPDITTVGVFHHSTGTMLYPABOIAEVVRLPDKA  
KIIDISAFGGICDYEKACDVLVAAPNKLKSGVPGISIIILAEAKGCARSA  
TLDLSQMSFDRKSGQFAVTPVHVYHALQALVEYKRDGSLSGKRTYQAKQALVRE  
AVKAMGTFTLGESKPCANVVCVNMPPDRPNWFKREYTHLNEGLIYPCAKASHAE  
TFREGIITHTSLANCERLMKCAKALKSMGIDHLQKSNM\*  
4506..5678  
/gene="L7535.2"  
4506..5678  
/gene="L7535.2"

/note="BLASTP similarity to met-10+ related protein from  
Methanobacterium thermoautotrophicum, and hypothetical  
proteins from several other organisms; COG similarity to  
SAM-dependent methyltransferases"

/codon\_start=1  
/product="L952.4"  
/protein\_id="AAF28368.1"  
/db\_xref="GI:6778464"

/translation="MLPATLRLLLSVSAALVLYVRKHFTPTPKGAKKLRGKEPAG  
RKPKGVAPQKIDITFTNVCAAFPEMSDVAVRRLFPKFEVHGHHVIVRLNDGTVEE  
LRPLARPAESFAPVLDVLDVGIIGELRRPSQILFOADTALTAYATSLRRTVQ  
RMRGARKGQPSGCMSTDAVESTLNRTASPTTAHVENGVIYSPDVSVMFSSGNTT  
ERIHFGVTADEVVDFNFCGIGYFTLPLAMHGNVAAIHALEKKNVSDIFVKLNALN  
KYDHLHPVCGDNREVEGELGKCDRVLWGIYIPTKSLPRAASFLLKNEAGRSSGVV  
HYHFLADKLCAAEALRDVQDELGEAAVFAFVRIADRCVKYAPARFHFVADLVFE\*  
6542..9043  
/gene="L7535.3"

6542..9043  
/gene="L7535.3"  
/note="TBLASTN similarity to Trypanosoma brucei genomic  
clones 28K20.TR (AQ648995) and 16113 (AQ657925) and EST  
clone ETN318 (AAB25012)."

/codon\_start=1  
/product="L7535.3"  
/protein\_id="AAF28369.1"  
/db\_xref="GI:6778465"

/translation="MEQELSKIIITEAARCSQGRDSETTAQVAFILSDAKGRNHVY  
QCCDLAODPGRRPRIMFLVNVQSPFROCLSEIGKFAHSGSPVSDQLKLAVG  
ALLAPLEGLSSSTARSROGYAEFALLADFSAPSGSGGGVPSISAVPYMKMG  
VSKKNEGVKCYSEVYQVVMCPYVDFQFLLCLESTVERQASINISANTYKTPNV  
MOTLFFMKQLCDGVYKLVSEPOLHFSRVTICYLPMLLEMESDYLCCVRNHVYDFI  
ENLGVHVLIDIPAGVYPCGTEALEOEVMVMAITVROAVLKICDETAAWAAKCLVA  
VYPCYRHADLRFVLOLPLGLMELHPEAFTALQAFAFARSLRNKDVDAADNLV  
VDELPAKLGSSAMSDIITIRCHTTPGARKALFOLLVACARROGVGARGSVMSG  
QLSPARAGAFRLMSDFFWYVFPLLIYMTSETVOQELPRRIADGLVSKDFGTKKS  
NVPLVLVENMIGLAEADELESYQAFKAEVLRREDTAFALQALLSEVAGTLP  
LILLEGREASDTKMTAAWRLAFVLSRMSARYLPDDAHKALTDLKRLNLFYDKGRH  
IGNRARVGMOCASLLNLLCLTSPQCDATFTVLEVLVDFDRGDAMDRTYKALYH  
HLIEYLCVPTRSSPHLNTANDVSDISAMIRERAMCINSAVOLLVGLVLMGMYRSLD  
OSOVAACVRAHRLVLLCHOSTDRRGVEVTRTTLTDPYPPVALLAAEKTIHIFKA  
BEKPTAVTEINCEAFRDYHDMRILEKGRKRY"

9610..11700  
/gene="CTP1"  
/note="L7535.4; L4625.4; predicted using Glimmer, TestCode  
and CodonUsage:ORF starts at 10883, but 2nd ATG is  
predicted as start codon by Glimmer, TestCode and  
CodonUsage"

9610..11700  
/gene="CTP1"  
/note="BLASTP similarity to CTR1 repressor from  
Saccharomyces cerevisiae (P32784), contains  
acyltransferase motifs and heme-binding motif"

/codon\_start=1  
/product="Choline Transporter Protein 1"  
/protein\_id="AAF28370.1"  
/db\_xref="GI:6778466"

/translation="MAESESTLSGVTAALTAAGASISIPGDPANAAETSPATSVPLDAS  
VSSAAEAPSAACSTPLTAAASPMAPTSAKPKDRIVASTRRKFLVCECHQNVPT  
IEWPDHRDLRLKSVLSEVSRFHHTLVRLMGFLVLRVLSYFREVTVGVRENIPCT  
GAVFYGNHONQFIDALMMHSHCGRPVRFMAEKSFORPVIQTLGRIFNSVPPVVRPD  
VPLVDECKLVTEGKFLVAGCTHFSRLSKGDLVSGVPGNACRAQVSRITSDSEVE  
VTVPIPAHEVHTPTSFYKSVRIDHSEMYAQVYDLQNHGICIGIPGEGSHDHTSILP  
LKAGVALFSLGAARHLSVKIVPVLGLTLYLGHKFSRATYIEGFAFPSPDDDLVKLFTD  
DKRAGTGLFLQLNAALRAVTINVPDYKTLTSLFSLFROLYQPPNCSMLSGHLLRLIR

gene  
CDS  
LGNVIEBQKSAEFAEFREKVENYSDFCALMIRDSOATLKRLLNADNEALQVGLLL  
RRFYALYLMVIFVFFVGLPIGGIIKYALAKQTKRALSSESSKIVGADVTGSFVL  
FAFVPGVFLVTFVFLVSDRTALVILSLPMMYVSLLIQEAIMELRAALPLL  
MSLVQHMOPKKLYERREDLARQARLIVHKYDPLEEEIKVHGNGDSNDLNRPSL  
FSLRYNRRQATNS\*  
12965..14668  
/gene="DP58"  
12965..14668  
/gene="DP58"  
/note="BLASTP similarity to DNA primase large subunit from  
several organisms"  
/codon\_start=1  
/product="DNA primase large subunit p58"  
/protein\_id="AAE28371.1"  
/db\_xref="GI:6778467"  
/translation="MQAITASTPQQYVSGAGIEKPLGVTTSDADMTMYERKPHGNST  
LFEAMVAKRMFLAWDOQINSPOAKNFDVLDIAIRLPEERSSAATDCGRVKM  
ISLGYDTAAGGSDRSSRSTGSAAGAOASIVFAPEEDVTSHLLCRPAFCMSE  
RWRDLVTERVLLTARIKEMAKRPTFLVLDLKRNGLPCLATDKOLADPMLQEV  
EYRVAESARESEGRVENYIAPLSLATRLIKRKYVLCRAGAILFRDQGVFLTV  
FCARLNGLHNAVLRVKKOALBEETAKSTVMAMLDALFOQFISDPVDTLQEGVAGSV  
CARDVORLAQHPFPCMRAIDWHLRREGHLKHHGFMYGFLKISIGLSLEDSLELAT  
LMKVGGGSVEAKTAYGYNVRHNYMEGKMSYSSASCATILGLPPVVDQHDCHGC  
PFRFDGALRTMLGKETENPKGCDYPSVRPTPCDIEDIVSDSKAOKHYTRACYKEMA  
TPGARDTLFRSPYEYYSLEFETPSDGTESARSPVPGKRTSVVLNEDVLRFTS  
P\*  
15866..16627  
/gene="COX11"  
15866..16627  
/gene="COX11"  
/note="BLASTP similarity to COX11P from several organisms"  
/codon\_start=1  
/product="Cytochrome c oxidase assembly protein"  
/protein\_id="AAE28372.1"  
/db\_xref="GI:6778468"  
/translation="MFRACWLRYCAAGATGGGGVSGPPPRKDNAGSSVSEPOARGSS  
RRRYRISGVDGADNYDERGQFSAFLSGLIISLGTFLFVPLRYMYCAPTGADPK  
FYTPDAARDEQDKSPYKPKLLKRLFLSDVGMTPIAFVPLQREVEVILGAPALF  
YSYRNSRNTLLGVSSYTIAPPEVTYNLKIQCFCFEQRKPHVELMEMPVFFYIDRD  
FLDSMYNLDVEVTYFNLEKTKDYIFRPNAR"  
17889..20552  
/gene="L7535.7"  
/note="L7535.7; predicted using Glimmer, TestCode and  
CodonUsage: ORF starts at 19017, but 2nd ATG is predicted  
as start codon by Glimmer, TestCode and CodonUsage"  
17889..20552  
/gene="L7535.7"  
/note="TBLASTN similarity to T. brucei sheared DNA genomic  
clones 29122 (AQ661463) and 12E2 (AQ655534)"  
/codon\_start=1  
/product="L7535.7"  
/protein\_id="AAE28373.1"  
/db\_xref="GI:6778469"  
/translation="MKGMPTLVIGRKELFSFADRYSSRRLLSDYSLADGEVLCLFNL  
VFDLIRPAPQTHSSTORAHANEQLFRFAKLHIPLAFLDPAALRTSGVEGFST  
LVLFYFHHLSKPDFSAEALDAVEEVTTLQSTECIALLLGLALQWRAVPEPLAQ  
TLRAHPFHHTAAEAAQEEAAHYTSMFAARRTGSGANSMHLLGPATRAHRSRTEQH  
SGIGSPMRPASSPSSAASSTPGFGTADSPPPRPLRNAAATQSTRSRLPSS  
DDVDCADLCRLHQCSERRTGAECAGVSOPTAPNQRSSRRSSASRRSSRNHQS  
TMSALSCRSSTAMVDAQPCVPYHSASSVSAATGSRGSAVSLRAHDALEAALQS  
KEACESLQMOVAOLLTLANTRAADTASATNGSHVSRPRRAAARGDADGTS  
TARLQHYADISIELETLQRAYREAGDKRQAPSPAPAPPPSGATMDKAALE  
AETALADIVDDGETITVDVHEKANLHCLHLEHQEAPRNRKQMWLWSTVAAH  
HTEGRMAAAVLLRLWASATLPAAGAGAAGAAVANSANREACTAQRQSWAAPFSA  
ASPGSDLDTSDEGACVKAETALSAHAELEHAREHOLAAALQAAQANNAATVQR  
AVRERLWKSLSACVAAEQAYAITETCAEEVEERLQREHYGIVSTQOLVED  
GQKVATSTOHOETATGTPSCRSLOTLVAHQEERAEILROVRLHIALASLOQE  
KQPTACADVVKVPKASLSAYASTAVAPFESDDLQRKSAVADGGAAGTSTDRECK  
PSSATYSGPVRAHRLTSLQLRGEFEFS\*  
21314..24091  
/gene="LRRP2"  
21314..24091

Query Match 84.0%; Score 16.8; DB 3; Length 37641;  
Best Local Similarity 90.0%; Pred. No. 6.9e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 caggttcacgcctcagctcc 20  
||||||| ||||| |||||  
Db 27551 CAGGTTCCAGGCTCCGCTCC 27532  
RESULT 14  
AC005767/c 40339 bp DNA linear INV 26-JAN-2000  
LOCUS Leishmania major chromosome 3 clone L7535 strain Friedlin, complete  
DEFINITION sequence.  
ACCESSION AC005767  
VERSION GI:5923657  
KEYWORDS HTG.  
SOURCE Leishmania major.  
ORGANISM Leishmania major.  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
Leishmania.  
REFERENCE 1 (bases 1 to 40339)  
AUTHORS Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrock,M.,  
Cawthra,J., Marsolini,F., Sunkin,S. and Stuart,K.D.  
TITLE Direct Submission  
JOURNAL Submitted (03-OCT-1998) Seattle Biomedical Research Institution, 4  
REFERENCE 2 (bases 1 to 40339)  
AUTHORS Nickerson Street, Seattle, WA 98109-1651, USA  
REFERENCE 3 (bases 1 to 40339)  
AUTHORS Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrock,M.,  
Cawthra,J., Marsolini,F., Sunkin,S. and Stuart,K.D.  
TITLE Direct Submission  
JOURNAL Submitted (25-SEP-1999) Seattle Biomedical Research Institution, 4  
REFERENCE 4 (bases 1 to 40339)  
AUTHORS Nickerson Street, Seattle, WA 98109-1651, USA  
TITLE Direct Submission  
JOURNAL Submitted (26-JAN-2000) Seattle Biomedical Research Institute, 4  
COMMENT On Sep 25, 1999 this sequence version replaced gi:5801659.  
FEATURES  
Location/Qualifiers  
1..40339  
/organism="Leishmania major"  
/strain="Friedlin"  
/db\_xref="taxon:5664"  
/chromosome="3"  
/clone="L7535"  
<2..559  
/gene="D3PCDH"  
/note="L7535.1L; L952.3; predicted using Glimmer, TestCode  
and CodonUsage"  
<2..559  
/gene="D3PCDH"  
/note="Catalyses first step in L-serine biosynthesis;  
BLASTP and COG similarity to phosphoglycerate  
dehydrogenase from several organisms"  
/codon\_start=1  
/product="D-3-phosphoglycerate dehydrogenase"  
/protein\_id="AAE27940.1"  
/db\_xref="GI:6759103"  
/translation="IRLMKRGSLYLNASRGTVVDLEALAKRGLHAGIADIVPEE  
PGSNKLEHRTPLQISNVILTPHVGSTCEAIGVEGTALAKFVTSGITAGAVNF  
PELVRRPVDPRKPLTNVHANVPGALNEINKVAVDLCNMGQFLLSTSKAIGYLMIDV  
DKDVAVELRKRISALKYSIRTLIR"  
1325..2652  
/gene="2AEPAT"  
/note="L7535.1; L952.1; predicted using Glimmer, TestCode  
and Codon Usage"



## RESULT 15

AC102139  
 LOCUS Mus musculus clone RP23-286F21, DNA linear HTG 23-NOV-2001  
 DEFINITION Mus musculus clone RP23-286F21, LOW-PASS SEQUENCE SAMPLING.  
 AC102139  
 VERSION GI:17061225  
 KEYWORDS HTG; HTGS\_PHASE0.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 58564)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Mus musculus, clone RP23-286F21  
 Unpublished

2 (bases 1 to 58564)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,  
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
 Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
 Jones,C., Kanat,A., Karatas,A., Kells,C., Labocque,K.,  
 Lanazares,R., Landers,T., Lehotzky,J., Levine,R., Liu,G.,  
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
 McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,  
 Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,  
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,  
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strausman,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,  
 Vain,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zieloun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L18086

Center clone name: 286\_F\_21

-----

\* NOTE: This record contains 74 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will

\* be preserved. 679: contig of 679 bp in length  
 \* 1  
 \* 680 779: gap of 100 bp  
 \* 780 1491: contig of 712 bp in length  
 \* 1492 1591: gap of 100 bp  
 \* 1592 2295: contig of 704 bp in length  
 \* 2296 2395: gap of 100 bp  
 \* 2396 3099: contig of 704 bp in length

\* 3100 3199: gap of 100 bp  
 \* 3200 3903: contig of 704 bp in length  
 \* 3904 4003: gap of 100 bp  
 \* 4004 4709: contig of 706 bp in length  
 \* 4710 4809: gap of 100 bp  
 \* 4810 5490: contig of 681 bp in length  
 \* 5491 5590: gap of 100 bp  
 \* 5591 6281: contig of 691 bp in length  
 \* 6282 6381: gap of 100 bp  
 \* 6382 7083: contig of 702 bp in length  
 \* 7084 7183: gap of 100 bp  
 \* 7184 7888: contig of 705 bp in length  
 \* 7889 7988: gap of 100 bp  
 \* 7989 8691: contig of 703 bp in length  
 \* 8692 8791: gap of 100 bp  
 \* 8792 9485: contig of 694 bp in length  
 \* 9486 9585: gap of 100 bp  
 \* 9586 10284: contig of 699 bp in length  
 \* 10285 10384: gap of 100 bp  
 \* 10385 11080: contig of 696 bp in length  
 \* 11081 11180: gap of 100 bp  
 \* 11181 11858: contig of 678 bp in length  
 \* 11859 11958: gap of 100 bp  
 \* 11959 12643: contig of 685 bp in length  
 \* 12644 12743: gap of 100 bp  
 \* 12744 13435: contig of 692 bp in length  
 \* 13436 13535: gap of 100 bp  
 \* 13536 14229: contig of 694 bp in length  
 \* 14230 14329: gap of 100 bp  
 \* 14330 15029: contig of 700 bp in length  
 \* 15030 15129: gap of 100 bp  
 \* 15130 15825: contig of 696 bp in length  
 \* 15826 15925: gap of 100 bp  
 \* 15926 16609: contig of 684 bp in length  
 \* 16610 16709: gap of 100 bp  
 \* 16710 17392: contig of 683 bp in length  
 \* 17393 17492: gap of 100 bp  
 \* 17493 18192: contig of 700 bp in length  
 \* 18193 18292: gap of 100 bp  
 \* 18293 18967: contig of 675 bp in length  
 \* 18968 19067: gap of 100 bp  
 \* 19068 19760: contig of 693 bp in length  
 \* 19761 19860: gap of 100 bp  
 \* 19861 20566: contig of 706 bp in length  
 \* 20567 20666: gap of 100 bp  
 \* 20667 21366: contig of 700 bp in length  
 \* 21367 21466: gap of 100 bp  
 \* 21467 22143: contig of 677 bp in length  
 \* 22144 22243: gap of 100 bp  
 \* 22244 22944: contig of 701 bp in length  
 \* 22945 23044: gap of 100 bp  
 \* 23045 23727: contig of 683 bp in length  
 \* 23728 23827: gap of 100 bp  
 \* 23828 24515: contig of 688 bp in length  
 \* 24516 24615: gap of 100 bp  
 \* 24616 25321: contig of 706 bp in length  
 \* 25322 25421: gap of 100 bp  
 \* 25422 26118: contig of 697 bp in length  
 \* 26119 26218: gap of 100 bp  
 \* 26219 26906: contig of 688 bp in length  
 \* 26907 27006: gap of 100 bp  
 \* 27007 27699: contig of 693 bp in length  
 \* 27700 27799: gap of 100 bp  
 \* 27800 28478: contig of 679 bp in length  
 \* 28479 28578: gap of 100 bp  
 \* 28579 29278: contig of 700 bp in length  
 \* 29279 29378: gap of 100 bp  
 \* 29379 30077: contig of 699 bp in length  
 \* 30078 30177: gap of 100 bp  
 \* 30178 30847: contig of 670 bp in length  
 \* 30848 30947: gap of 100 bp  
 \* 30948 31642: contig of 695 bp in length  
 \* 31643 31742: gap of 100 bp





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 14:58:40 ; Search time 2878.96 Seconds  
(without alignments)  
138.107 Million cell updates/sec

Title: US-09-623-329-24

Perfect score: 19

Sequence: 1 agtgttggaactctgtgg 19

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	19	100.0	19	6	AX018465
c 2	18	94.7	17281	14	EBV
c 3	18	94.7	184113	14	HS4B958RAJ
c 4	17.4	91.6	737	3	AF220666
c 5	17.4	91.6	5128	10	MMHDAC1
c 6	17.4	91.6	91573	2	AC094991
c 7	17.4	91.6	169018	2	AC105737
c 8	17.4	91.6	239336	2	AL591478
c 9	16.4	86.3	2676	10	AF275549
c 10	16.4	86.3	36654	2	AC017551
c 11	16.4	86.3	63806	2	AC107856
c 12	16.4	86.3	67599	2	AC010710
c 13	16.4	86.3	109101	2	AC108674
c 14	16.4	86.3	118447	2	AC093508
c 15	16.4	86.3	147518	2	AC073121
c 16	16.4	86.3	152470	9	AL135923
c 17	16.4	86.3	153558	9	AL365500
c 18	16.4	86.3	162608	9	AC009478
c 19	16.4	86.3	164798	9	AC092598
c 20	16.4	86.3	165951	2	AC104677
c 21	16.4	86.3	169915	2	AC068801
c 22	16.4	86.3	171374	2	AC068111
c 23	16.4	86.3	173434	2	AC016142
c 24	16.4	86.3	177735	3	AC005714
c 25	16.4	86.3	182118	9	AL356287
c 26	16.4	86.3	184767	2	AC084363
c 27	16.4	86.3	186634	2	AC099711
c 28	16.4	86.3	191466	3	AC008348
c 29	16.4	86.3	191775	2	AC074358
c 30	16.4	86.3	200364	2	AC095691
c 31	16.4	86.3	204292	2	AP004071
c 32	16.4	86.3	206443	2	AL669814
c 33	16.4	86.3	216692	2	AC099472
c 34	16.4	86.3	219809	2	AF336379
c 35	16.4	86.3	223020	9	AL162424
c 36	16.4	86.3	223335	2	AL392187
c 37	16.4	86.3	290792	3	AE003457
c 38	16	84.2	97266	2	AF274573
c 39	16	84.2	119104	9	AL157788
c 40	16	84.2	142064	9	AC091193
c 41	16	84.2	149965	9	AC007718
c 42	16	84.2	153619	9	AC026239
c 43	16	84.2	173071	2	AC094125
c 44	16	84.2	174727	2	AP000902
c 45	16	84.2	177694	2	AC019358

#### ALIGNMENTS

RESULT	1	19 bp	DNA	linear	PAT 07-SEP-2000
AX018465	Sequence 24 from Patent WO9945155.				
LOCUS	AX018465				
DEFINITION	Sequence 24 from Patent WO9945155.				
ACCESSION	AX018465				
VERSION	AX018465.1				
KEYWORDS	GI:10042616				
SOURCE	Epstein-Barr virus.				
ORGANISM	Human herpesvirus 4				
	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;				
	Gammaherpesvirinae; Lymphocryptovirus.				
REFERENCE	1 (bases 1 to 19)				
AUTHORS	Middelcorp, J.M., Van Den Brule, A.J. and Vervoort, M.B.				
TITLE	Oligonucleotides for the amplification and detection of Epstein				
JOURNAL	bar virus (ebv) nucleic acid				
	Patent: WO 9945155-A 24 10-SEP-1999;				
	MIDDELDOEP JAAP MICHIEL (NL); AKZO NOBEL NV (NL); DEN BRULE				
	ADRIANUS JOHANNES CH (NL); VERVOORT MARCEL BARTOLINA HEND (NL)				
FEATURES	Location/Qualifiers				
source	1..19				
	/organism="Human herpesvirus 4"				
	/db_xref="taxon:10376"				
BASE COUNT	2 a 3 c 7 g 7 t				

ORIGIN		JOURNAL MEDLINE	The EMBO journal. 2 (8), 1331-1338 (1983)
Query Match	100.0%; Score 19; DB 6; Length 19;	PUBMED	20331131
Best Local Similarity	100.0%; Pred. No. 3.5;	REFERENCE	10872327
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		AUTHORS	8 (bases 142687 to 159853)
		TITLE	Bankier,A.T., Deininger,P.L., Farrell,P.J. and Barrell,B.G. Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8 Epstein-Barr virus
Qy 1 agtgttgacactctgtgg 19		JOURNAL MEDLINE	Epstein-Barr virus Molecular biology & medicine. 1 (1), 21-45 (1983)
Db 1 AGTGTGGCACCCTCTGTGG 19		PUBMED	85035713
		PUBMED	6092825
RESULT 2		REFERENCE	9 (bases 112620 to 125316)
EBV/c		AUTHORS	Sequin,C., Farrell,P.J. and Barrell,B.G.
LOCUS		TITLE	DNA sequence and transcription of the BamHI fragment B region of B95-8 Epstein-Barr virus
DEFINITION	EBV 172281 bp DNA circular VRL 20-SEP-1999	JOURNAL MEDLINE	Molecular biology & medicine. 1 (3), 369-392 (1983)
ACCESSION	Epstein-Barr virus (EBV) genome, strain B95-8.	PUBMED	85060424
VERSION	V01555 J02070 K01729 K01730 V01554 X00498 X00499 X00784	REFERENCE	6094953
KEYWORDS	V01555.1 GI:59074	AUTHORS	10 (bases 45415 to 52824)
SOURCE	DNA polymerase; EBNA; genome; ribonucleotide reductase; tandem repeat; terminal repeat.	TITLE	Jones,M.D., Foster,L., Sheedy,T. and Griffin,B.E. The EB virus genome in Daudi Burkitt's lymphoma cells has a deletion similar to that observed in a non-transforming strain (P3HR-1) of the virus
ORGANISM	Epstein-Barr virus.	JOURNAL MEDLINE	The EMBO journal. 3 (4), 813-821 (1984)
REFERENCE	Human herpesvirus 4	PUBMED	84207939
AUTHORS	Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae; Lymphocryptovirus.	REFERENCE	6327290
TITLE	1 (bases 1 to 172281)	AUTHORS	11 (bases 87650 to 92703)
JOURNAL	Arrand,J.R., Rymo,L., Walsh,J.E., Bjorck,E., Lindahl,T. and Griffin,B.E.	TITLE	Biggin,M., Farrell,P.J. and Barrell,B.G. Transcription and DNA sequence of the BamHI L fragment of B95-8 Epstein-Barr virus
MEDLINE	Molecular cloning of the complete Epstein-Barr virus genome as a set of overlapping restriction endonuclease fragments	JOURNAL MEDLINE	The EMBO journal. 3 (5), 1083-1090 (1984)
PUBMED	Nucleic acids research. 9 (13), 2999-3014 (1981)	PUBMED	84236104
REFERENCE	82014887	AUTHORS	6203743
AUTHORS	2 (bases 1 to 172281)	TITLE	12 (bases 7315 to 9312)
JOURNAL	Kozak,M.	JOURNAL MEDLINE	Yates,J., Warren,N., Reisman,D. and Sugden,B. A cis-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells
PUBMED	Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes	REFERENCE	84222045
AUTHORS	Nucleic acids research. 9 (20), 5233-5262 (1981)	TITLE	Proceedings of the National Academy of Sciences of the United States of America. 81 (12), 3806-3810 (1984)
JOURNAL	82059504	JOURNAL MEDLINE	13 (bases 76089 to 79808)
PUBMED	3 (bases 1 to 172281)	REFERENCE	6328526
AUTHORS	Deininger,P.L., Bankier,A., Farrell,P., Baer,R. and Barrell,B.	AUTHORS	Gibson,T., Stockwell,P., Ginsburg,M. and Barrell,B. Homology between two EBV early genes and HSV ribonucleotide reductase and 38K genes
TITLE	Sequence analysis and in vitro transcription of portions of the Epstein-Barr virus genome	JOURNAL MEDLINE	Nucleic acids research. 12 (12), 5087-5099 (1984)
JOURNAL	Journal of cellular biochemistry. 19 (3), 267-274 (1982)	PUBMED	84247360
MEDLINE	83109311	REFERENCE	14 (bases 1 to 172281)
REFERENCE	4 (bases 1 to 172281)	AUTHORS	Baer,R.J., Bankier,A.T., Biggin,M.D., Deininger,P.L., Farrell,P.J., Gibson,T.J., Hatfull,G.F., Hudson,G.S., Satchwell,S.C., Seguin,C., Tuffnell,P.S. and Barrell,B.G.
AUTHORS	Farrell,P.J., Deininger,P.L., Bankier,A. and Barrell,B.	TITLE	DNA sequence and expression of the B95-8 Epstein-Barr virus genome
TITLE	Homologous upstream sequences near Epstein-Barr virus promoters	JOURNAL MEDLINE	Nature. 310 (5974), 207-211 (1984)
JOURNAL	Proceedings of the National Academy of Sciences of the United States of America. 80 (6), 1565-1569 (1983)	REFERENCE	6087149
MEDLINE	83169725	AUTHORS	15 (bases 1 to 172281)
REFERENCE	5 (bases 45644 to 52450)	TITLE	Bodescot,M. and Perricaudet,M. Clustered alternative splice sites in Epstein-Barr virus RNAs
AUTHORS	Jeang,K.T. and Hayward,S.D.	JOURNAL MEDLINE	Nucleic acids research. 15 (14), 5887 (1987)
TITLE	Organization of the Epstein-Barr virus DNA molecule. III. Location of the P3HR-1 deletion junction and characterization of the NotI repeat units that form part of the template for an abundant 12-O-tetradecanoylphorbol-13-acetate-induced mRNA transcript	PUBMED	87289053
JOURNAL	Journal of virology. 48 (1), 135-148 (1983)	REFERENCE	3039467
MEDLINE	83294686	AUTHORS	16 (bases 1 to 172281)
PUBMED	6310141	TITLE	Laux,G., Perricaudet,M. and Farrell,P.J. A spliced Epstein-Barr virus gene expressed in immortalized lymphocytes is created by circularization of the linear viral genome
AUTHORS	6 (bases 159853 to 172281)	JOURNAL MEDLINE	The EMBO journal. 7 (3), 769-774 (1988)
TITLE	Bankier,A.T., Deininger,P.L., Satchwell,S.C., Baer,R., Farrell,P.J. and Barrell,B.G.	PUBMED	88283646
JOURNAL	DNA sequence analysis of the EcoRI Dhet fragment of B95-8 Epstein-Barr virus containing the terminal repeat sequences	REFERENCE	2840285
MEDLINE	Molecular biology & medicine. 1 (4), 425-445 (1983)	AUTHORS	17 (bases 1 to 172281)
PUBMED	85060428	TITLE	Hatfull,G.F., Barrell,B.G., Quinn,J. and McGeoch,D. Unpublished
REFERENCE	6094955		
AUTHORS	7 (bases 1 to 172281)		
TITLE	Farrell,P.J., Bankier,A., Sequin,C., Deininger,P. and Barrell,B.G. Latent and lytic cycle promoters of Epstein-Barr virus		



Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 gttgtggcactctgtgg 19  
|||||  
Db 177591 GTGTGGCACTCTGTGG 177574

RESULT 4  
AF220666/c  
LOCUS  
DEFINITION  
Plasmodium vivax isolate KPVD8 98-1 Duffy receptor binding domain  
gene, partial cds.  
ACCESSION  
AF220666  
VERSION  
AF220666.1 GI:5694953  
KEYWORDS  
SOURCE  
malaria parasite P. vivax.  
ORGANISM  
Plasmodium vivax

REFERENCE  
1 (bases 1 to 737)  
AUTHORS  
Lim,C. and Song,J.-W.  
TITLE  
The analysis of Plasmodium vivax Duffy receptor binding domain gene  
sequence from resurgent Korea isolates

JOURNAL  
Unpublished  
REFERENCE  
2 (bases 1 to 737)  
AUTHORS  
Lim,C. and Song,J.-W.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (04-JAN-2000) Clinical Pathology, Korea University Ansan  
Hospital, 516, Gojan Dong, Ansan City, Kyunggi 425-020, South Korea

FEATURES  
source  
1..737  
/organism="Plasmodium vivax"  
/isolate="KPVD8 98-1"  
/db\_xref="taxon:5855"  
/clone="10"  
/country="South Korea"  
<1..>737  
/codon\_start=1  
/product="Duffy receptor binding domain"  
/protein\_id="AAF25492.1"  
/db\_xref="GI:5694953"

CDS

VEGDLLKLNRYNKDECKDIWISLGDGDIIMCTDMEGIGYSEVENNLSIFCTG  
EKAQHKKQWNNESKAGIWTAMYSVKRLKGAFWICKINAVNIEPQYIRIRWNG  
RDYVSELTPEVKDKCKDKGKINIDKRVCKVPPCKACKSDQWITRKNQWDLVSN  
KFKSVKNAEYQVAGIVTPYDILKQEL"  
BASE COUNT 297 a 90 c 160 g 190 t  
ORIGIN

Query Match 91.6%; Score 17.4; DB 3; Length 737;  
Best Local Similarity 94.7%; Pred. No. 35;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 agtgtggcactctgtgg 19  
|||||  
Db 521 AGTTTGGCACTCTGTGG 503

RESULT 5  
MMHDAC1  
LOCUS  
DEFINITION  
Mus musculus histone deacetylase 3 (Hdac3) gene, promoter and exons  
1, 2 and 3.  
ACCESSION  
AF079309  
VERSION  
AF079309.1 GI:3676555  
KEYWORDS  
SEGMENT  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus

REFERENCE  
1 (bases 1 to 5128)  
AUTHORS  
Mahlknecht,U., Bucala,R. and Verdin,E.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
JOURNAL  
FEATURES  
source  
1..5128  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="18"  
/map="B3"  
promoter  
1..791  
/gene="Hdac3"  
exon  
792..868  
/gene="Hdac3"  
/number=1  
intron  
869..962  
/evidence=not\_experimental  
/gene="Hdac3"  
/number=1  
exon  
963..1045  
/gene="Hdac3"  
/number=2  
intron  
1046..3102  
/gene="Hdac3"  
/number=2  
exon  
3103..3245  
/gene="Hdac3"  
/number=3  
intron  
3246..>5128  
/gene="Hdac3"  
/number=3

BASE COUNT 1293 a 1072 c 1371 g 1378 t 14 others  
ORIGIN

Query Match 91.6%; Score 17.4; DB 10; Length 5128;  
Best Local Similarity 94.7%; Pred. No. 38;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 agtgtggcactctgtgg 19  
|||||  
Db 1049 AGTGTGGCACTCTGTAG 1067

RESULT 6  
AC094991/c  
LOCUS  
DEFINITION  
Rattus norvegicus clone CH230-6F1, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
52 unordered pieces.

ACCESSION  
AC094991  
VERSION  
AC094991.2 GI:17941793  
KEYWORDS  
HTG: HTGS\_PHASE1.  
SOURCE  
Norway rat.  
ORGANISM  
Rattus norvegicus

REFERENCE  
1 (bases 1 to 91573)  
AUTHORS  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbaria,J.,  
Benton,J., Blincke,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
Burckhardt,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,  
Cartier,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
Coyle,M.D., D'Amore,S.R., David,R., Davila,M.L., Davis,C.,  
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,  
 Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,  
 Elhaj C., Escotto, M., Falls, T., Ferraguto, D., Flag, N., Ford, J.,  
 Foster, P., Prant, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,  
 Garza, N., Gill, R., Correll, J.H., Guevara, W., Gunaratne, P., Hale, S.,  
 Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,  
 Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C.,  
 Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,  
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,  
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,  
 Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,  
 Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,  
 Louisgied, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,  
 Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M.,  
 Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mollabbat, K.,  
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenko, S.,  
 Ogulu, M., Okwundu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,  
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,  
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojebokan, I., Rolfe, M.,  
 Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N.,  
 Sisson, I., Sodergren, E., Sonaika, T., Sparks, A., Stanley, H.,  
 Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,  
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Teifrod, B., Thomas, N.,  
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R.,  
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
 Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S.,  
 Worley, K., Wu, C., Wu, Y., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gibbs, R.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

Direct Submission  
 2 (bases 1 to 91573)  
 Worley, K.C.

Submitted (15-Sep-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Dec 20, 2001 this sequence version replaced gi:15624828.

COMMENT

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GBXG  
 Center clone name: CH230-6F1  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329Pfirst call to  
 findPhrapList  
 Consensus quality: 64526 bases at least Q40  
 Consensus quality: 71555 bases at least Q30  
 Consensus quality: 77402 bases at least Q20  
 Estimated insert size: 60446; sum-of-contigs estimation  
 Quality coverage: 0x in Q20 bases; agarose-fp estimation  
 Quality coverage: 0.6x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see <http://www.hgsc.bcm.tmc.edu/docs/Genbankdraft.data.html>).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 52 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 1532: contig of 1532 bp in length  
 \* 1533 1632: gap of unknown length  
 \* 1633 4149: contig of 2517 bp in length  
 \* 4150 4249: gap of unknown length  
 \* 4250 7308: contig of 3059 bp in length  
 \* 7309 7408: gap of unknown length

7409 10027: contig of 2619 bp in length  
 10028 10127: gap of unknown length  
 10128 12614: contig of 2487 bp in length  
 12614 12714: gap of unknown length  
 12714 15375: contig of 2661 bp in length  
 15375 15475: gap of unknown length  
 15475 17110: contig of 1635 bp in length  
 17110 17210: gap of unknown length  
 17210 20086: contig of 2876 bp in length  
 20086 20186: gap of unknown length  
 20186 22227: contig of 2041 bp in length  
 22227 22327: gap of unknown length  
 22327 24481: contig of 2154 bp in length  
 24481 24581: gap of unknown length  
 24581 26551: contig of 2070 bp in length  
 26551 26751: gap of unknown length  
 26751 29355: contig of 2604 bp in length  
 29355 29455: gap of unknown length  
 29455 31496: contig of 2041 bp in length  
 31496 31596: gap of unknown length  
 31596 33014: contig of 1418 bp in length  
 33014 33114: gap of unknown length  
 33114 35026: contig of 1912 bp in length  
 35026 35126: gap of unknown length  
 35126 36849: contig of 1723 bp in length  
 36849 36949: gap of unknown length  
 36949 39151: contig of 2202 bp in length  
 39151 39251: gap of unknown length  
 39251 40922: contig of 1671 bp in length  
 40922 41022: gap of unknown length  
 41022 43078: contig of 2056 bp in length  
 43078 43178: gap of unknown length  
 43178 44813: contig of 1635 bp in length  
 44813 46117: contig of 1204 bp in length  
 46117 46217: gap of unknown length  
 46217 48234: contig of 2017 bp in length  
 48234 48334: gap of unknown length  
 48334 49363: contig of 1029 bp in length  
 49363 49463: gap of unknown length  
 49463 51466: contig of 2003 bp in length  
 51466 51566: gap of unknown length  
 51566 52823: contig of 1257 bp in length  
 52823 52923: gap of unknown length  
 52923 54400: contig of 1477 bp in length  
 54400 54500: gap of unknown length  
 54500 56407: contig of 1907 bp in length  
 56407 56507: gap of unknown length  
 56507 57801: contig of 1294 bp in length  
 57801 57901: gap of unknown length  
 57901 59324: contig of 1423 bp in length  
 59324 59424: gap of unknown length  
 59424 60878: contig of 1454 bp in length  
 60878 62594: contig of 1616 bp in length  
 62594 62694: gap of unknown length  
 62694 64255: contig of 1561 bp in length  
 64255 64355: gap of unknown length  
 64355 65454: contig of 1099 bp in length  
 65454 65554: gap of unknown length  
 65554 67123: contig of 1569 bp in length  
 67123 67223: gap of unknown length  
 67223 68450: contig of 1227 bp in length  
 68450 68550: gap of unknown length  
 68550 69950: contig of 1400 bp in length  
 69950 70050: gap of unknown length  
 70050 71414: contig of 1364 bp in length  
 71414 71514: gap of unknown length  
 71514 72856: contig of 1342 bp in length  
 72856 72956: gap of unknown length  
 72956 74220: contig of 1264 bp in length  
 74220 74320: gap of unknown length  
 74320 75643: contig of 1323 bp in length

\* 75644 75743: gap of unknown length  
 \* 75744 contig of 1043 bp in length  
 \* 76786: contig of unknown length  
 \* 76887: contig of 1407 bp in length  
 \* 78293: contig of unknown length  
 \* 78294: contig of 1130 bp in length  
 \* 78394: contig of unknown length  
 \* 79524: contig of 1497 bp in length  
 \* 79624: contig of unknown length  
 \* 81120: contig of 1064 bp in length  
 \* 81221: contig of unknown length  
 \* 82285: contig of 1203 bp in length  
 \* 82388: contig of unknown length  
 \* 83587: contig of 1027 bp in length  
 \* 83688: contig of unknown length  
 \* 84714: contig of 1795 bp in length  
 \* 84815: contig of unknown length  
 \* 86609: contig of 1532 bp in length  
 \* 86710: contig of unknown length  
 \* 88241: contig of 1021 bp in length  
 \* 88242: contig of unknown length  
 \* 88342: contig of unknown length  
 \* 89466: contig of 1004 bp in length  
 \* 89463: contig of unknown length  
 \* 90467: contig of 1007 bp in length.  
 \* 90567 91573: contig of 1007 bp in length.

FEATURES  
 source

Location/Qualifiers  
 1..91573  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 /clone="CH230-6F1"  
 BASE COUNT 26358 a 16924 c 16717 g 26326 t 5248 others  
 ORIGIN

Query Match 91.6%; Score 17.4; DB 2; Length 91573;  
 Best Local Similarity 94.7%; Pred. No. 42;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agtgttgccactctgtg 19  
 |||||  
 Db 84075 AGTGTGGCCTCTGAGG 84057

RESULT 7  
 AC105737/c  
 LOCUS  
 DEFINITION  
 AC105737  
 AC105737.1 GI:18092982  
 VERSION  
 HTG; HTGS\_PHASE2; HTGS\_DRAFT".  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Sus scrofa  
 Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
 1 (bases 1 to 169018)  
 Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,  
 Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,  
 Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,  
 Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q.,  
 Legaspi,R., Maduro,Q.L., Maduro,V.B., Masiello,C., Maskeri,B.,  
 Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,  
 Stantrilpop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W.,  
 Tsurgoun,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,  
 Young,A., Zhang,L.-H. and Green,E.D.  
 NISC Comparative Sequencing Initiative  
 Unpublished  
 2 (bases 1 to 169018)  
 Green,E.D.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Direct Submission  
 Submitted (09-JAN-2002) NIH Intramural Sequencing Center, 8717  
 Grovemont Circle, Gaithersburg, MD 20877, USA  
 ----- Genome Center  
 Center: NIH Intramural Sequencing Center

Center code: NISC  
 Web site: http://www.nisc.nih.gov  
 Contact: nisc.mouse@nih.gov  
 ----- Project Information  
 Center project name: akk  
 Center clone name: 228E03  
 ----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads  
 Chemistry: dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 168847 bases at least Q40  
 Consensus quality: 168983 bases at least Q30  
 Consensus quality: 169012 bases at least Q20  
 Insert size: 153000; agarose-fp  
 Insert size: 165000; pulse-field-gel  
 Insert size: 169018; sum-of-contigs  
 Quality coverage: 13.15x in Q20 bases; agarose-fp  
 Quality coverage: 12.20x in Q20 bases; pulse-field-gel  
 Quality coverage: 11.91x in Q20 bases; sum-of-contigs

-----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 1 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 169018: contig of 169018 bp in length.

FEATURES  
 source

Location/Qualifiers  
 1..169018  
 /organism="Sus scrofa"  
 /db\_xref="taxon:9823"  
 /clone="RP44-228E3"  
 /clone\_lib="RP44"  
 misc\_feature  
 1..169018  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:left"

BASE COUNT 51934 a 31945 c 31431 g 53708 t  
 ORIGIN

Query Match 91.6%; Score 17.4; DB 2; Length 169018;  
 Best Local Similarity 94.7%; Pred. No. 43;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agtgttgccactctgtg 19  
 |||||  
 Db 122796 ACTGTGTCCTCTGCTG 122778

RESULT 8  
 AL591478/c  
 LOCUS  
 DEFINITION  
 AL591478  
 Mus musculus chromosome 2 clone RP23-140D14, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, in unordered pieces.  
 AL591478  
 AL591478.10 GI:17977733  
 VERSION  
 HTG; HTGS\_PHASE1; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
 house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (sites)  
 Phillimore,B.  
 Direct Submission  
 Submitted (07-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
 On Dec 23, 2001 this sequence version replaced gi:17644296.  
 ----- Genome Center

Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
 ----- Project Information  
 Center project name: bm140014

----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: plasmid; L08752; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 98% of reads  
 Chemistry: Dye-primer Big Dye; 1% of reads  
 Consensus quality: 239104 bases at least Q40  
 Consensus quality: 239180 bases at least Q30  
 Consensus quality: 239220 bases at least Q20  
 Insert size: 239236; sum-of-contigs  
 Insert size: 244850; 9.3% error; agarose-fp  
 Quality coverage: 11.37x in Q20 bases; sum-of-contigs Quality  
 coverage: 11.37x in Q20 bases; agarose-fp  
 -----

\* NOTE: This is a 'working draft' sequence.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

## FEATURES

## source

Location/Qualifiers  
 1. .239336  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="2"

misc\_feature  
 1. .99906  
 /note="assembly\_fragment:00112"

misc\_feature  
 10007..239336  
 /note="assembly\_fragment:07931"  
 BASE COUNT 61714 a 56185 c 57028 g 64309 t 100 others  
 ORIGIN

Query Match 91.6%; Score 17.4; DB 2; Length 239336;  
 Best Local Similarity 94.7%; Pred. No. 43;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agtgtggcactctctgtg 19

1 |||||  
 Db 39186 ACTGTTGGCACCCTCTGTGG 39168

## RESULT 9

AF275549/c  
 LOCUS AF275549 2676 bp mRNA linear ROD 12-SEP-2001  
 DEFINITION Mus musculus ubiquitin-associated protein NAG20 mRNA, complete cds.  
 ACCESSION AF275549  
 VERSION AF275549.3 GI:14488396

## SOURCE

house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 2676)  
 Qian, J., Dong, L., Zhang, B., Wang, J., Zhou, M., Li, Z., Li, W., Li, X.  
 and Li, G.

## REFERENCE

Identification and digital expression analysis of murine UBAP1 gene  
 by means of EST database searching  
 Shengwu Huaxue Yu Shengwu Wuli Jinzhan (2001) In press  
 2 (bases 1 to 2676)

## AUTHORS

Qian, J., Yang, J., Zhang, X., Zhang, B., Wang, J., Zhou, M., Tang, K.,  
 Li, W., Zeng, Z., Zhao, X., Shen, S. and Li, G.

Isolation and characterization of a novel cDNA, UBAP1, derived from  
 the tumor suppressor locus in human chromosome 9p21-22  
 J. Cancer Res. Clin. Oncol. 127 (10), 613-618 (2001)

## MEDLINE

## PUBMED

21485291  
 11599797

REFERENCE 3 (bases 1 to 2676)

## AUTHORS

## TITLE

## JOURNAL

Qian, J., Yang, J. and Li, G.

Direct Submission

Submitted (05-JUN-2000) Laboratory of Molecular/Cellular Genetics,  
 Cancer Research Institute, 88 Xiangya Road, Changsha, Hunan 410078,  
 P.R. China

## REFERENCE

## AUTHORS

## JOURNAL

4 (bases 1 to 2676)

Qian, J., Yang, J. and Li, G.

Direct Submission

Submitted (16-OCT-2000) Laboratory of Molecular/Cellular Genetics,  
 Cancer Research Institute, 88 Xiangya Road, Changsha, Hunan 410078,  
 P.R. China

## REMARK

## REFERENCE

## AUTHORS

## JOURNAL

Sequence update by submitter

5 (bases 1 to 2676)

Qian, J., Yang, J. and Li, G.

Direct Submission

Submitted (19-JUN-2001) Laboratory of Molecular/Cellular Genetics,  
 Cancer Research Institute, 88 Xiangya Road, Changsha, Hunan 410078,  
 P.R. China

## REMARK

## COMMENT

## FEATURES

## source

Sequence update by submitter  
 On Jun 19, 2001 this sequence version replaced gi:10803747.

Location/Qualifiers

1. .2676

/organism="Mus musculus"

/db\_xref="taxon:10090"

203..1528

/codon\_start=1

/product="ubiquitin-associated protein NAG20"

/protein\_id="AAG0981\_3"

/db\_xref="GI:1448839"

/translation="MASKKLGTVDVHGTFSYLDVDFKIGDKFKTPAKVGLPIGFSLPD  
 CLQVYREMYDFSELEKKTIEWAEDIKLQEAQRAEAEAEAEKVNKSGPGGDSKY  
 SPFKTNTATMPPIPNILASLQHNHILTPRVSSATKQKVLSPHTKADFNPADE  
 CEEDPFDNLEKLTIDKEELRNILVGTGPIMAQLDSDNTARGSGAVLQDEVLASL  
 EQATLDKPLKHPNGFITLPOLGNCEKMSLSKSVLPPIPTVYVNIKLSLSPKLDSDS  
 NQKTVKLASTFHSTCLSGASRLSKPSTQSSSELNGDHTLGLSALNLSGTEVPT  
 LTSSQMSLSVSVTESSPPDPCPTVPLNFSVQVPMPSQCPQALELQALSPSE  
 RQCVETVVMGYSDCVLRMRKKGENTEQILDYLFAGQLCEKGFDLF"

BASE COUNT 709 a 657 c 655 g 655 t

## ORIGIN

Query Match 86.3%; Score 16.4; DB 10; Length 2676;  
 Best Local Similarity 94.4%; Pred. No. 1.5e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agtgtggcactctctgtg 18

1 |||||  
 Db 1206 AGTGTGGCACCCTCTGTG 1189

## RESULT 10

## AC017551

## LOCUS

## DEFINITION

AC017551 36654 bp DNA linear HTG 10-DEC-1999  
 Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered  
 pieces

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

AC017551.1 GI:6554445

HTG; HTGS\_PHASE2.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 36654)

Adams, M. and Venter, J. C.

Direct Submission

Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,

Rockville, MD, USA

This sequence was identified as CDW:10211225 by the submitter.

For more information on this record e-mail to [fly@celera.com](mailto:fly@celera.com).

\* NOTE: This is a 'working draft' sequence.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and



\* the accession number will be preserved.

FEATURES  
 Location/Qualifiers  
 source 1. 36654  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"

BASE COUNT 10925 a 7894 c 7741 g 10094 t  
 ORIGIN

Query Match 86.3%; Score 16.4; DB 2; Length 36654;  
 Best Local Similarity 94.4%; Pred. No. 1.6e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 gtgttgccactctctgtg 19  
 |||||  
 Db 32449 GTGTGGCACTCTCTGAGG 32466

RESULT 11  
 AC107856/c

LOCUS AC107856 63806 bp DNA linear HTG 24-JAN-2002  
 DEFINITION Mus musculus clone RP23-10E23, LOW-PASS SEQUENCE SAMPLING.

ACCESSION AC107856

VERSION AC107856.1 GI:18308559

KEYWORDS HTG: HTGS PHASE0.

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 63806)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Unpublished

2 (bases 1 to 63806)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
 Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,  
 Brown,A., Canarata,J., Campolano,A., Chang,J., Chazaro,B.,  
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
 Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-pierre,N.,  
 Hagos,B., Horton,L., Hulme,W., Iliiev,I., Johnson,R., Jones,C.,  
 Kanat,A., Karatas,A., Kellis,C., Lakocque,K., Lamazares,R.,  
 Landers,T., Lehoczy,J., Levine,K., Liu,G., MacLean,C.,  
 MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,  
 McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Milhova,T.,  
 Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,  
 Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
 Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
 Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S.,  
 Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
 Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,  
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
 Zainoun,J., Zemek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www.seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L20633

Center clone name: 10\_E\_23

-----

\* NOTE: This record contains 78 individual

\* sequencing reads that have not been assembled into

\* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1 747: contig of 747 bp in length  
 748 847: gap of 100 bp  
 848 1583: contig of 736 bp in length  
 1584 1683: gap of 100 bp  
 1684 2399: contig of 716 bp in length  
 2400 2499: gap of 100 bp  
 2500 3225: contig of 726 bp in length  
 3226 3325: gap of 100 bp  
 3326 4058: contig of 733 bp in length  
 4059 4158: gap of 100 bp  
 4159 4885: contig of 727 bp in length  
 4886 4985: gap of 100 bp  
 4986 5687: contig of 702 bp in length  
 5688 5787: gap of 100 bp  
 5788 6473: contig of 686 bp in length  
 6474 6573: gap of 100 bp  
 6574 7298: contig of 725 bp in length  
 7299 7398: gap of 100 bp  
 7399 8122: contig of 724 bp in length  
 8123 8222: gap of 100 bp  
 8223 8965: contig of 743 bp in length  
 8966 9065: gap of 100 bp  
 9066 9792: contig of 727 bp in length  
 9793 9892: gap of 100 bp  
 9893 10613: contig of 721 bp in length  
 10614 10713: gap of 100 bp  
 10714 11443: contig of 729 bp in length  
 11443 11543: gap of 100 bp  
 11543 12240: contig of 698 bp in length  
 12241 12340: gap of 100 bp  
 12341 13057: contig of 717 bp in length  
 13058 13157: gap of 100 bp  
 13158 13871: contig of 714 bp in length  
 13872 13971: gap of 100 bp  
 13972 14691: contig of 720 bp in length  
 14692 14791: gap of 100 bp  
 14792 15511: contig of 720 bp in length  
 15512 15611: gap of 100 bp  
 15612 16346: contig of 735 bp in length  
 16347 16446: gap of 100 bp  
 16447 17180: contig of 734 bp in length  
 17181 17280: gap of 100 bp  
 17281 17999: contig of 719 bp in length  
 18000 18099: gap of 100 bp  
 18100 18818: contig of 719 bp in length  
 18819 18918: gap of 100 bp  
 18919 19635: contig of 717 bp in length  
 19636 19735: gap of 100 bp  
 19736 20432: contig of 697 bp in length  
 20433 20532: gap of 100 bp  
 20533 21245: contig of 713 bp in length  
 21246 21345: gap of 100 bp  
 21346 22061: contig of 716 bp in length  
 22062 22161: gap of 100 bp  
 22162 22879: contig of 718 bp in length  
 22880 22979: gap of 100 bp  
 22980 23705: contig of 726 bp in length  
 23706 23805: gap of 100 bp  
 23806 24550: contig of 745 bp in length  
 24551 24650: gap of 100 bp  
 24651 25387: contig of 737 bp in length  
 25388 25487: gap of 100 bp  
 25488 26227: contig of 740 bp in length  
 26228 26327: gap of 100 bp

\* 26328 27019: contig of 692 bp in length  
\* 27020 27119: gap of 100 bp  
\* 27120 27842: contig of 723 bp in length  
\* 27843 27942: gap of 100 bp  
\* 27943 28651: contig of 709 bp in length  
\* 28652 28751: gap of 100 bp  
\* 28752 29441: contig of 690 bp in length  
\* 29442 29541: gap of 100 bp  
\* 29542 30278: contig of 737 bp in length  
\* 30279 30378: gap of 100 bp  
\* 30379 31106: contig of 728 bp in length  
\* 31107 31206: gap of 100 bp  
\* 31207 31938: contig of 732 bp in length  
\* 31939 32038: gap of 100 bp  
\* 32039 32772: contig of 734 bp in length  
\* 32773 32872: gap of 100 bp  
\* 32873 33577: contig of 705 bp in length  
\* 33578 33677: gap of 100 bp  
\* 33678 34385: contig of 708 bp in length  
\* 34386 34485: gap of 100 bp  
\* 34486 35181: contig of 696 bp in length  
\* 35182 35281: gap of 100 bp  
\* 35282 35965: contig of 684 bp in length  
\* 35966 36085: gap of 100 bp  
\* 36086 36793: contig of 728 bp in length  
\* 36794 36893: gap of 100 bp  
\* 36894 37615: contig of 722 bp in length  
\* 37616 37715: gap of 100 bp  
\* 37716 38452: contig of 737 bp in length  
\* 38453 38552: gap of 100 bp  
\* 38553 39277: contig of 725 bp in length  
\* 39278 39377: gap of 100 bp  
\* 39378 40073: contig of 696 bp in length  
\* 40074 40173: gap of 100 bp  
\* 40174 40865: contig of 692 bp in length  
\* 40866 40965: gap of 100 bp  
\* 40966 41658: contig of 693 bp in length  
\* 41659 41758: gap of 100 bp  
\* 41759 42495: contig of 737 bp in length  
\* 42496 42595: gap of 100 bp  
\* 42596 43323: contig of 728 bp in length  
\* 43324 43423: gap of 100 bp  
\* 43424 44153: contig of 730 bp in length  
\* 44154 44253: gap of 100 bp  
\* 44254 44994: contig of 741 bp in length  
\* 44995 45094: gap of 100 bp  
\* 45095 45804: contig of 710 bp in length  
\* 45805 45904: gap of 100 bp  
\* 45905 46624: contig of 720 bp in length  
\* 46625 46724: gap of 100 bp  
\* 46725 47433: contig of 709 bp in length  
\* 47434 47533: gap of 100 bp  
\* 47534 48246: contig of 713 bp in length  
\* 48247 48346: gap of 100 bp  
\* 48347 49080: contig of 734 bp in length  
\* 49081 49180: gap of 100 bp  
\* 49181 49901: contig of 721 bp in length  
\* 49902 50001: gap of 100 bp  
\* 50002 50695: contig of 694 bp in length  
\* 50696 50795: gap of 100 bp  
\* 50796 51503: contig of 708 bp in length  
\* 51504 51603: gap of 100 bp  
\* 51604 52339: contig of 736 bp in length  
\* 52340 52439: gap of 100 bp  
\* 52440 53157: contig of 718 bp in length  
\* 53158 53257: gap of 100 bp  
\* 53258 53990: contig of 733 bp in length  
\* 53991 54090: gap of 100 bp  
\* 54091 54821: contig of 731 bp in length  
\* 54822 54921: gap of 100 bp  
\* 54922 55593: contig of 672 bp in length  
\* 55594 55693: gap of 100 bp  
\* 55694 56405: contig of 712 bp in length

\* 56406 56505: gap of 100 bp

Query Match 86.3%; Score 16.4; DB 2; Length 63806;

Best Local Similarity 94.4%; Pred. No. 1.6e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agtgtggcactctgtg 18

|||||||

DB 62975 AGTGTGGCATTCTGTG 62958

RESULT 12

AC010710/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

REFERENCE

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

AC010710 67599 bp DNA linear HTG 16-OCT-1999  
Drosophila melanogaster chromosome 3L/7 clone RCI98-27N2, \*\*\*  
SEQUENCING IN PROGRESS \*\*\*; 51 unordered pieces.

AC010710.2 GI:6056135

HTG: HTGS\_PHASE1.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 67599)

Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K.,  
Bodota, B., Bouck, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C.,  
Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,  
David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,  
Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,  
Forcum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L.,  
Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hoques, M.,  
Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M.,  
Kelly, S., Kondrjewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,  
Lichtarge, O., Liu, J., Liu, W., Logan, O., Lu, J., Lucier, R.,  
Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M., Morris, S.,  
Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S., Oswal, G.,  
Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L., Quiles, M.,  
Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S., Shah, E.,  
Shen, H., Simon, M., Sparks, A., Stamps, A., Sugeng, R., Tabor, P.,  
Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wahbah, M., Watlington, S.,  
Weinstock, G., Weinstock, I.R., Williamson, A., Worley, K., Wren, J.,  
Wrensford, G., Yu, W., Zhou, X., Nelson, D. and Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 67599)

Worley, K.C.

Direct Submission

Submitted (21-SEP-1999) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Oct 16, 1999 this sequence version replaced gi:5916429.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 51 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1112: contig of 1112 bp in length

\* 1113

\* 2277: contig of 1165 bp in length

\* 2278

\* 3154: contig of 877 bp in length

\* 3155

\* 4189: contig of 1035 bp in length

\* 4190

\* 5015: contig of 826 bp in length

\* 5016

\* 5857: contig of 842 bp in length

\* 5858

\* 7086: contig of 1229 bp in length

\* 7087

\* 7771: contig of 685 bp in length

\* 7772

\* 8972: contig of 1201 bp in length

\* 8973

\* 10048: contig of 1076 bp in length

\* 10049

\* 10954: contig of 906 bp in length

\* 10955

\* 11812: contig of 858 bp in length

\* 11813

\* 12984: contig of 1172 bp in length

```

* 12985 14078: contig of 1094 bp in length
* 14849: contig of 771 bp in length
* 14850 15785: contig of 936 bp in length
* 15786 16669: contig of 884 bp in length
* 16670 18287: contig of 1618 bp in length
* 18288 19623: contig of 1336 bp in length
* 20696: contig of 1073 bp in length
* 20697 21321: contig of 625 bp in length
* 21322 22676: contig of 1355 bp in length
* 22677 24400: contig of 1724 bp in length
* 24401 25650: contig of 1250 bp in length
* 25651 25522: contig of 872 bp in length
* 25523 27688: contig of 1163 bp in length
* 27689 28535: contig of 850 bp in length
* 28536 30156: contig of 1621 bp in length
* 30157 31435: contig of 1279 bp in length
* 31436 33344: contig of 1909 bp in length
* 33345 35180: contig of 1836 bp in length
* 35181 36650: contig of 1469 bp in length
* 36651 37605: contig of 955 bp in length
* 37606 39011: contig of 1407 bp in length
* 39012 40135: contig of 1124 bp in length
* 40136 41664: contig of 1529 bp in length
* 41665 43219: contig of 1555 bp in length
* 43220 44646: contig of 1427 bp in length
* 44647 46249: contig of 1603 bp in length
* 46250 47644: contig of 1395 bp in length
* 47645 49369: contig of 1725 bp in length
* 49370 50533: contig of 1164 bp in length
* 50534 51995: contig of 1462 bp in length
* 51996 53847: contig of 1852 bp in length
* 53848 55097: contig of 1250 bp in length
* 55098 57159: contig of 2062 bp in length
* 57160 58312: contig of 1153 bp in length
* 58313 60121: contig of 1809 bp in length
* 60122 61910: contig of 1789 bp in length
* 61911 64010: contig of 2100 bp in length
* 64011 67599: contig of 3589 bp in length.

FEATURES             Location/Qualifiers
     source            1..67599
                        /organism="Drosophila melanogaster"
                        /db_xref="taxon:7227"
                        /chromosome="3L/?"
                        /clone="RPC198-27N2"
BASE COUNT      17901 a 15671 c 15122 g 18857 t      48 others
ORIGIN

Query Match      86.3%; Score 16.4; DB 2; length 67599;
Best Local Similarity 94.4%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gttgtgacctctgtgg 19
    |||||
Db 3588 GTGTTGGCACTCTGAGG 3571

RESULT 13
AC108674/c
LOCUS
DEFINITION
Homo sapiens chromosome 3q clone RP11-37318, *** SEQUENCING IN
PROGRESS ***, 30 unordered pieces.
ACCESSION
AC108674.1 GI:18449709
VERSION
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 109101)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaral, H.C., Are, J.R., Ayale, M., Banks, T.,
Barbata, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,

```

```

Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Foster, P., Frantz, P.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harri, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B.,
Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S.,
Karlssoon, E., Kelly, S., Khan, O., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, M., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,
Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 109101)
Worley, K.C.
Direct Submission
Submitted (31-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: naylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HCMB
Center clone name: RP11-37318
----- Summary Statistics
Sequencing vector: Plasmid: M77789
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 113583 bases at least Q40
Consensus quality: 124622 bases at least Q30
Consensus quality: 132555 bases at least Q20
Estimated insert size: 118297; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 1.6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

```

\* as soon as it is available and the accession number will  
\* be preserved.

1 9692: contig of 9692 bp in length  
\* 9693: gap of unknown length  
\* 9793: contig of 6925 bp in length  
\* 16717: gap of unknown length  
\* 16817: gap of unknown length  
\* 22054: contig of 5237 bp in length  
\* 22154: gap of unknown length  
\* 22155: contig of 5923 bp in length  
\* 28078: gap of unknown length  
\* 28177: gap of unknown length  
\* 34638: contig of 6461 bp in length  
\* 34738: gap of unknown length  
\* 34639: contig of 4816 bp in length  
\* 34739: gap of unknown length  
\* 39555: contig of 4404 bp in length  
\* 39658: gap of unknown length  
\* 44058: contig of 4404 bp in length  
\* 44159: gap of unknown length  
\* 44159: contig of 4485 bp in length  
\* 48644: gap of unknown length  
\* 48744: contig of 3383 bp in length  
\* 52127: gap of unknown length  
\* 52227: contig of 3231 bp in length  
\* 55457: gap of unknown length  
\* 55557: gap of unknown length  
\* 55558: contig of 2897 bp in length  
\* 58455: gap of unknown length  
\* 58455: contig of 3579 bp in length  
\* 62133: gap of unknown length  
\* 62134: gap of unknown length  
\* 62234: contig of 3150 bp in length  
\* 65384: gap of unknown length  
\* 65484: contig of 2177 bp in length  
\* 67660: gap of unknown length  
\* 67661: contig of 3434 bp in length  
\* 71194: gap of unknown length  
\* 71294: gap of unknown length  
\* 74410: contig of 3116 bp in length  
\* 74411: gap of unknown length  
\* 74511: contig of 2763 bp in length  
\* 77273: gap of unknown length  
\* 77374: contig of 2076 bp in length  
\* 79449: gap of unknown length  
\* 79450: contig of 2111 bp in length  
\* 79550: gap of unknown length  
\* 81661: gap of unknown length  
\* 81761: contig of 2391 bp in length  
\* 84251: gap of unknown length  
\* 84252: contig of 3151 bp in length  
\* 87402: gap of unknown length  
\* 87502: contig of 2213 bp in length  
\* 87503: gap of unknown length  
\* 89716: gap of unknown length  
\* 89716: contig of 2103 bp in length  
\* 89816: gap of unknown length  
\* 91918: contig of 2763 bp in length  
\* 92018: gap of unknown length  
\* 94781: contig of 2109 bp in length  
\* 94782: gap of unknown length  
\* 94881: contig of 2109 bp in length  
\* 96990: gap of unknown length  
\* 97090: contig of 2303 bp in length  
\* 99393: gap of unknown length  
\* 99394: contig of 2150 bp in length  
\* 99494: gap of unknown length  
\* 101644: contig of 2005 bp in length  
\* 101744: gap of unknown length  
\* 103748: gap of unknown length  
\* 103749: contig of 2839 bp in length  
\* 103848: gap of unknown length  
\* 106687: gap of unknown length  
\* 106688: contig of 2314 bp in length  
\* 106788: contig of 2314 bp in length.

FEATURES  
source

1. .109101  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3q"  
/clone="RP11-37318"

BASE COUNT 30913 a 21874 c 22391 g 30962 t 2961 others  
ORIGIN

Query Match  
Best Local Similarity

86.3% Score 16.4; DB 2; Length 109101;  
94.4% Pred. No. 1.7e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 gtgtggcactctgtgg 19  
||||| |||||  
Db 27352 GTGTGGCGCTCTGTGG 27335

RESULT 14  
AC093508/c

LOCUS 118447 bp DNA linear HTG 31-AUG-2001  
DEFINITION Homo sapiens chromosome 16 clone CTD-2255G7, WORKING DRAFT  
SEQUENCE, 6 unordered pieces.

AC093508  
AC093508.1 GI:15383797  
HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 118447)  
DOE Joint Genome Institute.

AUTHORS

Sequencing of Human Chromosome 16  
Unpublished

JOURNAL

2 (bases 1 to 118447)  
DOE Joint Genome Institute.

REFERENCE

Direct Submission  
DOE Joint Genome Institute.

AUTHORS

Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

JOURNAL

-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov

COMMENT

Project Information  
Center Project Name: 725620  
Center clone name: CTRB-HI\_2255G7

Summary Statistics

Consensus quality: 112694 bases at least Q40  
Consensus quality: 115488 bases at least Q30  
Consensus quality: 115940 bases at least Q20  
Estimated insert size: 120000; agarose-fp estimation  
Quality coverage: 8.1 in Q20 bases; agarose-fp estimation  
Quality coverage: 8.24 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 6 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1002: contig of 1002 bp in length  
\* 1003: gap of unknown length  
\* 1103: contig of 1030 bp in length  
\* 2133: gap of unknown length  
\* 2233: contig of 1032 bp in length  
\* 3265: gap of unknown length  
\* 3364: contig of 10779 bp in length  
\* 3365: gap of unknown length  
\* 14144: contig of 30599 bp in length  
\* 14244: gap of unknown length  
\* 44843: contig of 73505 bp in length.  
\* 44943: 118447: contig of 73505 bp in length.

FEATURES  
source

1. .118447  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="CTD-2255G7"

BASE COUNT 29465 a 28364 c 29105 g 31009 t 504 others  
ORIGIN

Query Match 86.3%; Score 16.4; DB 2; Length 118447;  
 Best Local Similarity 94.4%; Pred. No. 1.7e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 gtgttgccactctgtgg 19  
 ||| ||||| ||||| |||||  
 Db 86559 GTGGTGGCACTTCTGTGG 86542

FEATURES  
 source

1. 147518  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /clone="RP11-612F4"

misc\_feature 1..1242  
 /note="assembly\_name:Contig17"  
 misc\_feature 1343..2720  
 /note="assembly\_name:Contig20"

misc\_feature 2821..4024  
 /note="assembly\_name:Contig22"  
 misc\_feature 4125..5475  
 /note="assembly\_name:Contig26"

misc\_feature 5576..7112  
 /note="assembly\_name:Contig30"  
 misc\_feature 7213..27350  
 /note="assembly\_name:Contig32"

clone\_end:SP6  
 vector\_side:left  
 misc\_feature 27451..28202  
 /note="assembly\_name:Contig24"

misc\_feature 28303..82991  
 /note="assembly\_name:Contig33"  
 misc\_feature 83092..147518  
 /note="assembly\_name:Contig34"

BASE COUNT 46898 a 25346 c 25595 g 48869 t 810 others  
 ORIGIN

Query Match 86.3%; Score 16.4; DB 2; Length 147518;  
 Best Local Similarity 94.4%; Pred. No. 1.7e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gtgttgccactctgtgg 19  
 ||||| ||||| ||||| |||||

Db 132026 GTGGTGGCACTTCTGTGG 132043

Search completed: May 18, 2002, 14:59:01  
 Job time: 8238 sec

RESULT 15  
 AC073121  
 LOCUS Homo sapiens chromosome 2 clone RP11-612F4, WORKING DRAFT SEQUENCE,  
 9 unordered pieces.  
 AC073121  
 AC073121.3 GI:13027578  
 HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 147518)  
 Waterston,R.H.  
 The sequence of Homo sapiens clone  
 Unpublished

2 (bases 1 to 147518)  
 Waterston,R.H.  
 Direct Submission  
 Submitted (08-JUN-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA

COMMENT On Feb 21, 2001 this sequence version replaced gi:9665215.  
 ----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 ----- Project Information -----  
 Center project name: H.NH0612F04  
 ----- Summary Statistics -----  
 Sequencing vector: M13; 86%  
 Sequencing vector: plasmid; 10%  
 Chemistry: Dye-primer ET; 86% of reads  
 Assembly: Dye-terminator Big Dye; 10% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 142201 bases at least Q40  
 Consensus quality: 143721 bases at least Q40  
 Consensus quality: 144493 bases at least Q20  
 Insert size: 145000; agarose-fp  
 Quality coverage: 5.53 in Q20 bases; sum-of-contigs  
 Quality coverage: 5.56 in Q20 bases; sum-of-contigs  
 -----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 9 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1242: contig of 1242 bp in length  
 \* 1243 1342: gap of unknown length  
 \* 1343 2720: contig of 1378 bp in length  
 \* 2721 2820: gap of unknown length  
 \* 2821 4024: contig of 1204 bp in length  
 \* 4025 4124: gap of unknown length  
 \* 4125 5475: contig of 1351 bp in length  
 \* 5476 5576: gap of unknown length  
 \* 5576 7112: contig of 1537 bp in length  
 \* 7113 7212: gap of unknown length

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:45:56 ; Search time 1024.22 Seconds  
(without alignments)  
33.526 Million cell updates/sec

Title: US-09-623-329-23  
Perfect score: 20  
Sequence: 1 ggctgtcaccgtttcttgg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

- 1: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDSS/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	AAZ11673	Oligo specific for
2	16.8	84.0	414	AA188789	Human polynucleoti
3	16.8	84.0	3375	AAQ56053	DPM2 mannosyl tran
C 4	16.4	82.0	276	AA187648	Human polynucleoti
C 5	16.4	82.0	548	AA113031	Human breast cance
C 6	16.4	82.0	678	AA121901	Human breast cance
C 7	16.4	82.0	918	AAH03852	Human cDNA clone (
8	16.4	82.0	1092	AAH03852	Arabidopsis thalia
9	16.4	82.0	1094	AAC33824	Arabidopsis thalia

C 10	16.4	82.0	1729	22	AAH13851	Human cDNA sequenc
C 11	16.4	82.0	2997	16	AAQ99209	Rat metabotropic g
C 12	15.8	79.0	372	22	ABA08614	Human NHP2 protein
C 13	15.8	79.0	712	21	AAC49491	Arabidopsis thalia
C 14	15.8	79.0	817	22	AAF44800	Cysteine protease
C 15	15.8	79.0	900	22	AAQ50308	Mouse alpha-1,3 ga
C 16	15.8	79.0	1203	21	AAC81128	Human secreted pro
C 17	15.8	79.0	1404	20	AAQ90185	Mouse presenilin-1
C 18	15.8	79.0	1964	17	AAQ40030	Mouse presenilin-1
C 19	15.8	79.0	1964	19	AAV04668	Mouse presenilin-1
C 20	15.8	79.0	2681	18	AAT64819	Tumour suppressor
C 21	15.8	79.0	5721	22	AA104026	Human reproductive
C 22	15.8	79.0	9431	22	AA104905	Human reproductive
C 23	15.8	79.0	9434	22	AA104904	Human reproductive
C 24	15.8	79.0	48974	20	AAQ55300	Mouse Presenilin-1
C 25	15.4	77.0	268	20	AAV89908	EST clone CT489.
C 26	15.4	77.0	270	22	ABA89171	Escherichia coli p
C 27	15.4	77.0	1017	22	AAF33067	Human secreted pro
C 28	15.4	77.0	1557	21	AAQ54069	Brain specific mem
C 29	15.4	77.0	1570	21	AAC76369	Human ORFX ORF1924
C 30	15.4	77.0	1573	24	AAQ56233	CDNA sequence #20
C 31	15.4	77.0	1575	22	AAH99723	Human protein enco
C 32	15.4	77.0	1581	21	AAZ54236	Neisseria meningit
C 33	15.4	77.0	1581	21	AAZ54238	Neisseria meningit
C 34	15.4	77.0	1602	22	AAQ52428	Human secreted pro
C 35	15.4	77.0	1628	20	AAQ60585	Human secreted pro
C 36	15.4	77.0	1901	15	AAQ66178	Seven transmembran
C 37	15.4	77.0	1901	19	AAV18356	Human R12 seven tr
C 38	15.4	77.0	1901	21	AAQ91725	Human 7TM receptor
C 39	15.4	77.0	3077	22	AAQ55223	Human secreted pro
C 40	15.4	77.0	3861	11	AAQ05979	Recombinant AcNPV-
C 41	15.4	77.0	4113	11	AAQ06644	Recombinant AcNPV-
C 42	15.4	77.0	12259	22	AAQ36190	Human cardiovasscul
C 43	15.4	77.0	19199	22	AAK70995	Human immune/haema
C 44	15.4	77.0	48354	22	ABA89141	Escherichia coli p
C 45	15.4	77.0	48345	22	ABA89142	Escherichia coli p

## ALIGNMENTS

RESULT 1  
AAZ11673  
ID AAZ11673 standard; DNA; 20 BP.  
XX  
AC AAZ11673;  
XX  
DT 19-NOV-1999 (first entry)  
XX  
DE Oligo specific for EBV BARF-1 RNA.

XX Epstein Barr Virus; EBV infection; viral; gene transcription; EBER-1;  
XX Epstein Barr early RNA; Epstein Barr nuclear antigen 1; EBNA-1; LMP-1;  
KW Latent membrane protein; LMP-2; vIL10; BCRF-1; BARF1; BDLF2; NASBA;  
KW EBV-associated malignancy; primer; ss.  
XX  
OS Synthetic.  
OS Epstein-barr virus.  
XX  
PN WO9945155-A2.  
XX  
PD 10-SEP-1999.  
XX  
PF 01-MAR-1999; 99WO-EP01392.  
XX  
PR 04-MAR-1998; 98EP-0200655.  
XX  
PR 14-DEC-1998; 98EP-0204231.  
XX  
PA (ALKU ) AKZO NOBEL NV.  
XX  
XX Vervoort MBHJ, Van Den Brule AJC, Middeldorp JM;  
XX WPI; 1999-551051/46.

XX Identifying Epstein Barr Virus Infection -  
 XX Claim 11; Page 20; 50pp; English.  
 XX The invention provides methods for identifying an Epstein Barr Virus  
 CC (EBV) infection, that comprises determining viral gene transcription  
 CC patterns by amplification of specific RNA sequences. The binding sites  
 CC of the oligos suitable for amplification are located in the following  
 CC genes: Epstein Barr early RNA (EBER-1), Epstein Barr nuclear antigen 1  
 CC (EBNA-1), latent membrane protein 1 (LMP-1), LMP-2, and vIL10 (BCRF-1),  
 CC BARF1 and BDLF2. The method comprises (a) amplifying a target sequence  
 CC within one or more RNA(s) transcribed from above gene sequences and the  
 CC (b) detecting the amplified products, determining the transcription  
 CC pattern and identifying the corresponding EBV-associated malignancy. The  
 CC RNA is amplified using a transcription based amplification technique  
 CC such as NASBA. The invention is used to diagnose malignant and  
 CC non-malignant EBV-associated diseases. Sequences AA211672-75 represent  
 CC oligos specific for BARF-1 RNA.  
 XX  
 SQ Sequence 20 BP; 1 A; 6 C; 6 G; 7 T; 0 other;  
 Query Match 100.0%; Score 20; DB 20; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 1.7;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 ggctgtcaccgctttcttgg 20  
 Db 1 ggctgtcaccgctttcttgg 20  
 RESULT 2  
 AA188789  
 ID AA188789 standard; cDNA; 414 BP.  
 XX  
 AC AA188789;  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 8849.  
 XX  
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200164835-A2.  
 XX  
 PD 07-SEP-2001.  
 XX  
 PF 26-FEB-2001; 2001WO-US04927.  
 XX  
 PR 28-FEB-2000; 2000US-0515126.  
 PR 18-MAY-2000; 2000US-0577409.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX  
 DR WPI; 2001-514838/56.  
 DR P-PSDB; AA008858.  
 XX  
 PT Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -  
 XX  
 PS Claim 1; SEQ ID NO 8849; 1399pp + Sequence Listing; English.  
 XX  
 CC The invention relates to human polynucleotides (AA179941-AA193841) and  
 CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to

CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 414 BP; 79 A; 126 C; 108 G; 100 T; 1 other;  
 Query Match 84.0%; Score 16.8; DB 22; Length 414;  
 Best Local Similarity 90.0%; Pred. No. 81;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 ggctgtcaccgctttcttgg 20  
 Db 323 gctgtcaccgctttcttgg 342  
 RESULT 3  
 AAQ56053  
 ID AAQ56053 standard; cDNA; 3375 BP.  
 XX  
 AC AAQ56053;  
 DT 12-SEP-1994 (first entry)  
 XX  
 DE DPM2 mannosyl transferase gene.  
 XX  
 KW Fungi; Saccharomyces cerevisiae; mannosyl transferase;  
 KW glycosylation; DPM2; fermentation; pharmacokinetics;  
 KW antigen reactions; ss.  
 XX  
 OS Saccharomyces cerevisiae.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 532..2985  
 FT /\*tag= a  
 FT /product= Dol-P-Man mannosyl transferase (DPM2).  
 PN DE4226971-A.  
 XX  
 XX 17-FEB-1994.  
 PD  
 PF 14-AUG-1992; 92DE-4226971.  
 XX  
 PR 14-AUG-1992; 92DE-4226971.  
 XX  
 PA (STRA/) STRAHL-BOLSINGER S.  
 PA (TANN/) TANNER W.  
 XX  
 PI Strahl-bolsinger S, Tanner W;  
 XX  
 DR WPI; 1994-058257/08.  
 DR P-PSDB; AAR47201.  
 XX  
 PT Fungal cells contg. mutated DPM2 mannosyl transferase gene - for  
 PT reduced O-glycosylation of prods., useful in pharmaceutical  
 PT prods. with altered pharmacokinetic features.  
 XX  
 PS Example 5; Figure 4a; 22pp; German.  
 XX  
 CC The Dol-P-Man:protein (Ser/Thr) mannosyl transferase (DPM2) is an  
 CC enzyme involved in the O-glycosylation cycle. Mutants of the DPM2  
 CC result in reduced levels of O-glycosylation. Undesirable  
 CC O-glycosylation of recombinant products results in antigen reactions  
 CC and different pharmacokinetic features e.g. different in vivo

CC clearance rate and plasma half life. Fungal cells transformed with  
 CC genes encoding mutant DPM2 are still viable and can be used in the  
 CC pharmaceutical industry for the production of recombinant products  
 CC without any undesirable O-glycosylation occurring.

XX  
 SQ Sequence 3375 BP; 893 A; 704 C; 706 G; 1072 T; 0 other;

Query Match 84.0%; Score 16.8; DB 15; Length 3375;  
 Best Local Similarity 90.0%; Pred. No. 1e+02; Mismatches 2; Indels 0; Gaps 0;  
 Matches 18; Conservative 0;

Qy 1 ggctgtcacgcgtttcttg 20  
 |||||  
 Db 2304 ggctgtcacgcgtttctgcg 2323

RESULT 4  
 AAI87648/c  
 ID AAI87648 standard; cDNA; 276 BP.

XX AC AAI87648;

XX DT 06-NOV-2001 (first entry)

XX DE Human polynucleotide SEQ ID NO 7708.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation; ss.

XX OS Homo sapiens.

XX PN WO200164835-A2.

XX PD 07-SEP-2001.

XX PF 26-FEB-2001; 2001WO-US04927.

XX PR 28-FEB-2000; 2000US-0515126.

XX PR 18-MAY-2000; 2000US-0577409.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX PR WPI; 2001-511838/56.

XX DR P-PSDB; AA007717.

XX PT Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -

XX PS Claim 1; SEQ ID NO 7708; 1399pp + Sequence Listing; English.

XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO0010-AAO13910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 276 BP; 67 A; 79 C; 82 G; 48 T; 0 other;

Query Match 82.0%; Score 16.4; DB 22; Length 276;  
 Best Local Similarity 94.4%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ctgtcacgcgtttcttg 20  
 |||||  
 Db 148 CTGTCACCGCTTCTTG 131

RESULT 5  
 AAL13031/c  
 ID AAL13031 standard; cDNA; 548 BP.

XX AC AAL13031;

XX DT 07-DEC-2001 (first entry)

XX DE Human breast cancer expressed polynucleotide 5488.

XX KW Human; breast cancer; cell marker; cytostatic; ss.

XX OS Homo sapiens.

XX PN WO200151628-A2.

XX PD 19-JUL-2001.

XX PF 10-JAN-2001; 2001WO-US00798.

XX PR 14-JAN-2000; 2000US-0176077.

XX PR 14-MAR-2000; 2000US-0189167.

XX PR 24-MAR-2000; 2000US-0192099.

XX PR 29-MAR-2000; 2000US-0193480.

XX PR 15-MAY-2000; 2000US-0205230.

XX PR 09-JUN-2000; 2000US-0211315.

XX PR 25-JUL-2000; 2000US-0220534.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Lillie J, Xu Y, Wang Y, Steinmann K;

XX PR WPI; 2001-451856/48.

XX PT New peptide useful as a marker for the diagnosis of breast cancer -

XX PS Claim 1; Page 986; 3695pp; English.

XX CC The invention relates to human breast cancer expressed polynucleotides  
 CC (AAL07544-AAL26789) and methods of assessing whether a patient is  
 CC afflicted with breast cancer by examining the correlation between the  
 CC expression of certain markers and the cancerous state of breast cells.  
 CC The polynucleotides and encoded polypeptides are potential markers for  
 CC detecting, diagnosing, monitoring, characterising treating and  
 CC potentially preventing breast cancer. The polynucleotides and encoded  
 CC polypeptides are also useful for isolating compounds with cytostatic  
 CC activity.

XX SQ Sequence 548 BP; 172 A; 100 C; 140 G; 136 T; 0 other;

Query Match 82.0%; Score 16.4; DB 22; Length 548;  
 Best Local Similarity 94.4%; Pred. No. 1.3e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 gctgtcacgcgtttcttg 19  
 |||||  
 Db 80 GCTGTACCGCTTCTTG 63

RESULT 6  
 AAL21901/c  
 ID AAL21901 standard; cDNA; 678 BP.



```
XX AC AAL21901;
XX PF 07-DEC-2001 (first entry)
XX DT Human breast cancer expressed polynucleotide 14358.
XX DE Human; breast cancer; cell marker; cytostatic; ss.
XX KW Homo sapiens.
XX OS WO200151628-A2.
XX PN 19-JUL-2001.
XX PD 10-JAN-2001; 2001WO-US00798.
XX PF 14-JAN-2000; 2000US-0176077.
XX PR 14-MAR-2000; 2000US-0189167.
XX PR 24-MAR-2000; 2000US-0192099.
XX PR 29-MAR-2000; 2000US-0193460.
XX PR 15-MAY-2000; 2000US-0205230.
XX PR 09-JUN-2000; 2000US-0211315.
XX PR 25-JUL-2000; 2000US-0220534.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX WPI; 2001-451856/48.
XX DR New peptide useful as a marker for the diagnosis of breast cancer
XX PT Claim 1; Page 2574; 3695pp; English.
XX PS The invention relates to human breast cancer expressed polynucleotides
XX CC (AAL07544-AAL26789) and methods of assessing whether a patient is
XX CC afflicted with breast cancer by examining the correlation between the
XX CC expression of certain markers and the cancerous state of breast cells.
XX CC The polynucleotides and encoded polypeptides are potential markers for
XX CC detecting, diagnosing, monitoring, characterising treating and
XX CC potentially preventing breast cancer. The polynucleotides and encoded
XX CC polypeptides are also useful for isolating compounds with cytostatic
XX CC activity.
XX CC Sequence 678 BP; 219 A; 120 C; 168 G; 171 T; 0 other;
XX SQ

Query Match 82.0%; Score 16.4; DB 22; Length 678;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 gctgtcaccgctttcttg 19
Db 58 GCTGTCAGCGCTTCTTG 41

RESULT 7
AAH03852/c
ID AAH03852 standard; cDNA; 918 BP.
XX AC AAH03852;
XX DT 26-JUN-2001 (first entry)
XX DE Human cDNA clone (5'-primer) SEQ ID NO:687.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX OS Homo sapiens.
XX KW Ep1074617-A2.
XX PN

PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PL Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX DR Primer sets for synthesizing polynucleotides, particularly the 5602
XX PT full-length cDNAs defined in the specification, and for the detection
XX PR and/or diagnosis of the abnormality of the proteins encoded by the
XX PT full-length cDNAs -
XX CC Claim 1; SEQ ID 687; 2537pp + CD ROM; English.
XX PS The present invention describes primer sets for synthesising 5602
XX CC full-length cDNAs defined in the specification. Where a primer set
XX CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX CC to the complementary strand of a polynucleotide which comprises one of
XX CC the 5602 nucleotide sequences defined in the specification, where the
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX CC of an oligonucleotide comprising a sequence complementary to the
XX CC complementary strand of a polynucleotide which comprises a 5'-end
XX CC sequence and an oligonucleotide comprising a sequence complementary to a
XX CC polynucleotide which comprises a 3'-end sequence, where the
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in
XX CC the specification. The primer sets can be used in antisense therapy and
XX CC in gene therapy. The primers are useful for synthesising polynucleotides,
XX CC particularly full-length cDNAs. The primers are also useful for the
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by
XX CC the full-length cDNAs. The primers allow obtaining of the full-length
XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
XX CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX CC represent oligonucleotides, all of which are used in the exemplification
XX CC of the present invention.
XX SQ Sequence 918 BP; 296 A; 175 C; 225 G; 220 T; 2 other;

Query Match 82.0%; Score 16.4; DB 22; Length 918;
Best Local Similarity 94.4%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 gctgtcaccgctttcttg 19
Db 271 GCTGTCAGCGCTTCTTG 254

RESULT 8
AAC50896
ID AAC50896 standard; DNA; 1092 BP.
XX AC AAC50896;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 66534.
XX KW Hybridisation assay; genetic mapping; gene expression control;
XX KW protein identification; signal transduction pathway;
XX KW metabolic pathway; promoter; termination sequence; ss.
XX PN
```

OS Arabidopsis thaliana.  
XX EPI033405-A2.  
XX  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 21-JUN-1999; 99US-0139859.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 12-AUG-1999; 99US-0148319.  
PR 13-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.

PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 82.0%; Score 16.4; DB 21; Length 1092;  
Best Local Similarity 94.4%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ctgtcacgcgttcttgg 20  
||| ||||| ||||| |||||  
Db 846 ctgcacgcgttcttgg 863

RESULT 9  
AAC33824  
ID AAC33824 standard; DNA; 1094 BP.

XX AAC33824;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 4445.

XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

XX EP1033405-A2.

PN 06-SEP-2000.

PD 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0121800.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126284.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 30-JUN-1999; 99US-0140991.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.

```
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161362.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      82.0%; Score 16.4; DB 21; Length 1094;
Best Local Similarity 94.4%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ctgtcacgcgtttcttgg 20
   ||| ||||| ||||| |||||
Db 848 ctgtcacgcgtttcttgg 865

RESULT 10
AAH13851/C
ID AAH13851 standard; cDNA; 1729 BP.
XX
AC AAH13851;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:10832.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
```

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 DR WPI: 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 10832; 2537pp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesising 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesising polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 1729 BP; 568 A; 300 C; 341 G; 520 T; 0 other;

Query Match 82.0%; Score 16.4; DB 22; Length 1729;  
 Best Local Similarity 94.4%; Pred. No. 1.5e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gctgtaccgcgtttcttg 19  
 ||||| ||||| |||||  
 Db 271 GCTGTACGCGTCTTCTG 254

RESULT 11  
 AAQ99209/c  
 ID AAQ99209 standard; cDNA; 2997 BP.  
 XX  
 AC AAQ99209;  
 XX  
 DT 23-JAN-1996 (first entry)  
 XX  
 DE Rat metabotropic glutamate receptor mGluR7.  
 XX  
 KW Metabotropic glutamate receptor; mGluR7; cerebral ischaemia;  
 KW migraine; Parkinson's disease; Huntington's chorea; ss.  
 XX  
 OS Rattus rattus.  
 XX  
 FH Key Location/Qualifiers  
 FT cds 133..2877  
 FT /\*tag= a  
 FT signal\_peptide 135..227  
 FT /\*tag= b  
 XX  
 PN W09518154-A1.  
 XX  
 PD 06-JUL-1995.  
 XX  
 PF 29-DEC-1994; 94WO-US14989.

XX 30-DEC-1993; 93US-0176401.  
 XX (UYOR-) UNIV OREGON HEALTH SCI.  
 PA (ZYMO) ZYMOGENETICS INC.  
 XX  
 PI Kinzie MJ, Mulvihill ER, Saugstad JA, Segerson TP;  
 PI Westbrook GL;  
 XX  
 DR WPI: 1995-246338/32.  
 DR P-PSDB; AAR80479.  
 XX  
 PT New metabotropic glutamate receptor mGluR7 and related antibodies  
 PT useful for identifying its modulators, diagnosis and treatment of  
 PT cerebral ischaemia etc.  
 XX  
 PS Claim 12; Page 25; 41pp; English.  
 XX  
 CC The mGluR7 protein encoded by this DNA sequence can bind glutamate  
 CC and induce cytoplasmic signal transduction. It is especially  
 CC involved in regulation of mitral/tufted neurons in the olfactory  
 CC bulb and may act by mediating inhibition of transmitter release at  
 CC particular glutamatergic synapses.  
 XX  
 SQ Sequence 2997 BP; 770 A; 789 C; 751 G; 687 T; 0 other;

Query Match 82.0%; Score 16.4; DB 16; Length 2997;  
 Best Local Similarity 94.4%; Pred. No. 1.5e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gctgtaccgcgtttcttg 19  
 ||||| ||||| |||||  
 Db 2210 GCTGTACCGATTCTTG 2193

RESULT 12  
 ABA08614/c  
 ID ABA08614 standard; cDNA; 372 BP.  
 XX  
 AC ABA08614;  
 XX  
 DT 11-JAN-2002 (first entry)  
 XX  
 DE Human NRP2 protein homologue-encoding cDNA, SEQ ID NO:390.  
 XX  
 KW Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;  
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
 KW chronic inflammatory condition; proliferative retinopathy;  
 KW atherosclerosis; coronary heart disease; arterial ischaemia;  
 KW bone disorder; osteoporosis; vascular growth disorder;  
 KW tissue regeneration; wound healing; infection; immune disorder;  
 KW cell culture; drug screening; gene therapy; antiinflammatory;  
 KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
 KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
 KW antifungal; vulnery; antiulcer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200157188-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WO-US03800.  
 XX  
 PR 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX

PI Tang YT, Liu C, Drmanac RT;  
XX MPI: 2001-457740/49.  
DR P-PSDB; ABB11370.  
XX Human proteins and DNA encoding sequences useful for preventing,  
PT treating or ameliorating a medical condition in a mammalian subject  
PT e.g. arthritis and cancer.  
XX Claim 1; Page 479; 1963pp; English.  
XX  
XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
CC invention also relates to vectors and recombinant host cells comprising a  
CC nucleotide of the invention, methods of producing the novel polypeptides,  
CC antibodies against the polypeptides, methods of detecting the nucleotides  
CC or polypeptides in a sample, and methods of identifying compounds which  
CC bind to polypeptides of the invention. Although novel, many of the  
CC polypeptides of the invention have homology to known proteins, thereby  
CC giving an insight into their probable biological activities, and hence  
CC potential therapeutic applications. The polypeptides of the invention may  
CC have various activities, including cytokine, cell proliferation or cell  
CC differentiation activities; stem cell growth factor activity;  
CC haematopoiesis regulatory activity; tissue growth activity;  
CC immunomodulatory activity; activin- or inhibin-related activities;  
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
CC thrombolytic activities; receptor or ligand activities; or may be  
CC involved in oncogenesis, cancer cell proliferation or metastasis.  
CC Depending on their biological activities, polypeptides and nucleotides of  
CC the invention are useful for preventing, treating or ameliorating medical  
CC conditions, e.g., by protein or gene therapy. Such conditions include  
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
CC vascular growth. Polypeptides involved with tissue regeneration and  
CC repair (or nucleic acids encoding them) may be used to promote wound  
CC healing (e.g., of burns, incisions and ulcers), while those with  
CC immunomodulatory activities may be used in the treatment of viral,  
CC bacterial and fungal infections in addition to immune disorders.  
CC Polypeptides with growth factor activity may be used in cell cultures to  
CC promote cell growth. For example, such polypeptides may be used to  
CC manipulate stem cells in culture to give rise to neuroepithelial cells  
CC that can be used to augment or replace cells damaged by illness,  
CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
CC may also be used in the diagnosis of the above conditions, and in drug  
CC screening techniques. The present sequence represents a cDNA encoding a  
CC novel human polypeptide of the invention.  
XX  
XX Sequence 372 BP; 99 A; 98 C; 111 G; 64 T; 0 other;  
SQ

Query Match 79.08; Score 15.8; DB 22; Length 372;  
Best Local Similarity 89.5%; Pred. No. 2.5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 ggctgtaccgcttcttg 19  
||| |||||  
Db 182 GGGTTCCACCGCTTCGTG 164

RESULT 13  
AAC49491  
ID AAC49491 standard; DNA; 712 BP.  
XX  
XX AAC49491;  
AC  
XX 18-OCT-2000 (first entry)  
DT  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 61354.  
DE  
XX Hybridisation assay; genetic mapping; gene expression control;  
XX protein identification; signal transduction pathway;  
KW

KW metabolic pathway; promoter; termination sequence; ss.  
XX Arabidopsis thaliana.  
OS  
XX EP1033405-A2.  
PN  
XX 06-SEP-2000.  
PD  
XX 25-FEB-2000; 2000EP-0301439.  
PF  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142220.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145313.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149502.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.

PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 79.0%; Score 15.8; DB 21; Length 712;  
Best Local Similarity 89.5%; Pred. No. 2.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ggctgtcacgcgtttcttg 19  
||||||| || |||||  
Db 613 ggctgtcacgcgtttcttg 631

## RESULT 14

AAF4800/C  
ID AAF4800 standard; cDNA; 817 BP.

XX AAF4800;

XX 27-MAR-2001 (first entry)

XX Cysteine protease coding sequence #14.

XX Cell death modulator; programmed cell death; PCD; apoptosis;  
XX forestry plant; ss.

XX Eucalyptus grandis.

XX WO200075331-A1.

XX 14-DEC-2000.

XX

```

PF 02-JUN-2000; 2000WO-NZ00086.
XX
XX
PR 04-JUN-1999; 99US-0325932.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Flinn B, Lasham A;
PI
PI WPI: 2001-061724/07.
DR P-PSDB; AAB65773.
XX
XX Novel defender against cell death polynucleotide useful for modulating
PT programmed cell death pathway and specific development pathways in
PT forestry plant.
XX
XX Claim 1; Page 89; 142pp; English.
XX
XX The present invention relates to coding sequences (see AAF44740-F44840
CC and AAF4843-F4844) and proteins (see AAB65714-B65814) involved in
CC programmed cell death (PCD; apoptosis). The coding sequences and proteins
CC of the present invention are useful for modulating a PCD or cell death
CC pathway and various developmental pathways in a forestry plant, by
CC stably incorporating one of the present coding sequences into the genome
CC of the forestry plant, where the coding sequence provides a PCD pathway
CC that is not present in a native form of the forestry plant.
XX
XX Sequence 817 BP; 212 A; 215 C; 223 G; 167 T; 0 other;
SQ
Query Match 79.0%; Score 15.8; DB 22; Length 817;
Best Local Similarity 89.5%; Pred. No. 2.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 gctgtcacgcgtcttctgg 20
I ||||| ||||| |||||
Db 464 GATGTCACAGCTTCTTGG 446

RESULT 15
AAS05308
ID AAS05308 standard; DNA; 900 BP.
XX
XX AAS05308;
AC
XX
DT 07-SEP-2001 (first entry)
XX
XX Mouse alpha-1,3 galactosyltransferase genomic sequence #3.
XX
XX Mouse; murine; alpha-1,3 galactosyltransferase promoter;
KW recombinant expression cassette; gene expression; transgenic animal;
KW livestock; growth hormone; xenotransplantation; transgenic cell; ds.
XX
XX Mus musculus.
XX
XX Key Location/Qualifiers
FH intron 1..336
FT /*tag= a
FT misc_feature 337..517
FT /*tag= b
FT /*note= "Untranslated exon 3"
FT intron 518..900
FT /*tag= c
XX
XX WO200130992-A2.
XX
XX 03-MAY-2001.
XX
XX 20-OCT-2000; 2000WO-US29139.
XX
XX 22-OCT-1999; 99US-0161092.
XX 25-AUG-2000; 2000US-0227951.
XX

```

---

```

PA (UYPI-) UNIV PITTSBURGH.
XX
XX Kolke C;
XX
XX WPI: 2001-300505/31.
XX
XX New recombinant expression cassette useful in xenotransplantation -
XX
XX Claim 16; Page 65-66; 138pp; English.
XX
XX The present sequence for mouse alpha-1,3 galactosyltransferase (GT)
CC genomic sequence #3 is 1 of 44 alpha-1,3 galactosyltransferase
CC gene sequences (AAS05290-AAS05333) from different animals. The
CC invention provides a novel recombinant expression cassette comprising
CC an alpha-1,3 GT promoter operably linked to a polynucleotide for
CC expression. The cassettes of the invention can be used to express
CC foreign genes or to disrupt the native alpha-1,3 GT genomic sequence.
CC The invention provides transgenic animals and methods for their
CC production. Transgenic mice in which a reporter gene is operably linked
CC to the alpha-1,3 GT promoter can be used for assessing promoter activity
CC and specificity. Transgenic livestock having the novel expression
CC cassette in which a growth hormone is expressed under the control of the
CC alpha-1,3 GT promoter can be matured or grown better than commonly
CC employed strains. Tissue obtained from transgenic animals can be
CC implanted into a host providing a method of xenotransplantation from a
CC transgenic animal. The invention also provides a transgenic organ
CC consisting essentially of transgenic cells engineered, for use in
CC transplantation. The methods of the invention facilitate
CC xenotransplantation between species, particularly between species
CC exhibiting differential expression of the gal-alpha-gal epitope. The
CC invention also describes primers (AAS05334-AAS05385) used to isolate the
CC alpha-1,3 GT sequences.
XX
XX Sequence 900 BP; 193 A; 245 C; 183 G; 269 T; 10 other;
SQ
Query Match 79.0%; Score 15.8; DB 22; Length 900;
Best Local Similarity 89.5%; Pred. No. 2.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ggcgtgcacgcgtcttcttg 19
I |||| ||||| ||||| |||
Db 639 ggctctcaccccttcttg 657

Search completed: May 18, 2002, 17:45:58
Job time: 13747 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:22:19 ; Search time 8624.33 Seconds  
(without alignments)  
29.735 Million cell updates/sec

Title: us-09-623-329-24  
Perfect score: 19  
Sequence: 1 agtgttgcaactctctgtg 19

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_hic:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_hic:\*
- 12: gb\_gss:\*
- 13: em\_gss\_hum:\*
- 14: em\_gss\_inv:\*
- 15: em\_gss\_pln:\*
- 16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.4	91.6	257	9	AW482287
2	17.4	91.6	411	9	AW447186
3	17.4	91.6	468	9	AW314573
C 4	17.4	91.6	495	10	BF711408
C 5	17.4	91.6	496	10	BF711205
C 6	17.4	91.6	504	10	BF712784
C 7	17.4	91.6	515	10	BM432649
C 8	17.4	91.6	522	10	BF704178
C 9	17.4	91.6	545	10	BM286313
10	17.4	91.6	551	10	BE683254
11	17.4	91.6	554	10	BE750214
12	17.4	91.6	556	10	BE683263
13	17.4	91.6	628	10	BF759242
C 14	17	89.5	159	12	AQ942809
C 15	17	89.5	567	9	AA694624
C 16	17	89.5	591	12	TA197E100
C 17	16.4	86.3	230	10	BJ017869

C 18	16.4	86.3	254	9	AW183265
C 19	16.4	86.3	443	12	AZ091464
C 20	16.4	86.3	452	10	BF022959
C 21	16.4	86.3	514	9	BB867852
C 22	16.4	86.3	543	12	AQ603208
C 23	16.4	86.3	560	12	AZ807079
C 24	16.4	86.3	598	10	BF178210
C 25	16.4	86.3	619	9	BB656545
C 26	16.4	86.3	680	12	AZ082327
C 27	16.4	86.3	681	12	AZ393232
C 28	16.4	86.3	709	9	BB615354
C 29	16.4	86.3	711	12	BH530414
C 30	16.4	86.3	770	10	BI734864
C 31	16.4	86.3	888	10	BF144533
C 32	16.4	86.3	965	12	CNS056H2
C 33	16.4	86.3	985	9	AI893934
C 34	16.4	86.3	990	9	AV132698
C 35	16.4	86.3	1059	12	CNS03LAR
C 36	16.4	86.3	2471	11	AK009596
C 37	16	84.2	386	10	BF424530
C 38	16	84.2	464	9	AW756024
C 39	16	84.2	694	12	AG125630
C 40	16	84.2	742	9	AI491491
C 41	16	84.2	760	10	BG914600
C 42	16	84.2	795	12	CNS056E7
C 43	16	84.2	820	10	BE311281
C 44	15.8	83.2	152	10	BG183551
C 45	15.8	83.2	165	10	BG208967

ALIGNMENTS

RESULT 1  
AW482287  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT

AW482287 257 bp mRNA linear EST 25-APR-2001  
41846 M46 3BOV Bos taurus cDNA 5', mRNA sequence.  
AW482287  
AW482287.1 GI:7052393  
EST.  
COW.  
Bos taurus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 257)  
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,  
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett  
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,  
Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and  
Keele,J.W.  
Sequence evaluation of four pooled-tissue normalized bovine cDNA  
libraries and construction of a gene index for cattle  
Genome Res. 11 (4), 626-630 (2001)  
21180013  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 20  
and -minmatch 12 options.  
PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCTCAGTCACGACG  
Plate: 21 row: M column: 3  
Seq primer: ATTTAGGTGACACTATAG.  
Location/Qualifiers  
1. .257  
/organism="Bos taurus"  
/db\_xref="taxon:9913"

```

/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."

BASE COUNT      75 a   33 c   47 g   102 t
ORIGIN

Query Match      91.6%; Score 17.4; DB 9; Length 257;
Best Local Similarity 94.7%; Pred. No. 3.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agtgtggcactctgtgg 19
|||||
Db 6 AGTGTGGCAGCTGTGTGG 24

RESULT 2
AW447186
LOCUS      88400 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION
ACCESSION  AW447186
VERSION     AW447186.1 GI:6988973
KEYWORDS   EST.
SOURCE     COW.
ORGANISM   Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 411)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCAGTCAGCAGC
Plate: 63 row: P column: 9
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers
1. .411
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."

BASE COUNT      147 a   64 c   73 g   127 t
ORIGIN

Query Match      91.6%; Score 17.4; DB 9; Length 411;
Best Local Similarity 94.7%; Pred. No. 3.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."

BASE COUNT      75 a   33 c   47 g   102 t
ORIGIN

Query Match      91.6%; Score 17.4; DB 9; Length 257;
Best Local Similarity 94.7%; Pred. No. 3.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agtgtggcactctgtgg 19
|||||
Db 6 AGTGTGGCAGCTGTGTGG 24

RESULT 2
AW447186
LOCUS      88400 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION
ACCESSION  AW447186
VERSION     AW447186.1 GI:6988973
KEYWORDS   EST.
SOURCE     COW.
ORGANISM   Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 411)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCAGTCAGCAGC
Plate: 63 row: P column: 9
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers
1. .411
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."

BASE COUNT      147 a   64 c   73 g   127 t
ORIGIN

Query Match      91.6%; Score 17.4; DB 9; Length 411;
Best Local Similarity 94.7%; Pred. No. 3.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 agtgtggcactctgtgg 19
|||||
Db 382 AGTGTGGCAGCTGTGTGG 400

RESULT 3
AW314573
LOCUS      10794 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION
ACCESSION  AW314573
VERSION     AW314573.1 GI:6743829
KEYWORDS   EST.
SOURCE     COW.
ORGANISM   Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 468)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCAGTCAGCAGC
Plate: 7 row: H column: 18
Seq primer: ATTAGTGACACTATAG.
Location/Qualifiers
1. .468
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 4BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 20 and day 40
embryos."

BASE COUNT      154 a   72 c   78 g   164 t
ORIGIN

Query Match      91.6%; Score 17.4; DB 9; Length 468;
Best Local Similarity 94.7%; Pred. No. 3.8e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agtgtggcactctgtgg 19
|||||
Db 306 AGTGTGGCAGCTGTGTGG 324

RESULT 4
BF711408/c
LOCUS      495 bp mRNA linear EST 02-JAN-2001
DEFINITION  MI-P-A2-acp-c-07-1-UM.s1 MI-P-A2 Sus scrofa clone
ACCESSION  BF711408
VERSION     BF711408.1 GI:12010883
KEYWORDS   EST.

```

SOURCE ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
REFERENCE 1 (bases 1 to 495)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
COMMENT Contact: Tuggle CK  
Molecular Genetics Laboratory, Department of Animal Science  
Iowa State University  
201 Kildee Hall, Ames, IA 50011-3150, USA  
Tel: 5152944252  
Fax: 5152942401  
Email: cktuggle@iastate.edu  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to verify it as a clone from the  
non-normalized anterior pituitary at estrus day 5 library cDNA  
Library Preparation: BJ Woods, JA Green, RS Prather S142 Animal  
Science Research Center, Department of Animal Science, University  
of Missouri-Columbia, 65211 Clone distribution: clones will be  
available through Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA=Yes.

FEATURES Location/Qualifiers  
source 1..495  
/organism="Sus scrofa"  
/strain="crossbreed"  
/db\_xref="taxon:9823"  
/clone="MI-P-A2-acp-c-07-1-UM"  
/clone\_lib="MI-P-A2"  
/lab\_host="DH10B (Life Technologies)"  
/note="vector: p7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: EcoRI; The MI-P-A2  
library is derived from anterior pituitary at estrus day  
5. For a detailed description of the library from which  
this clone was derived, please visit our web site at  
http://pigest.genome.iastate.edu/  
TAG\_LIB=MI-P-A2  
TAG\_TISSUE=anterior pituitary at estrus day 5  
TAG\_SEQ=TCGGCT"  
BASE COUNT 188 a 86 c 71 g 150 t  
ORIGIN  
Query Match 91.6%; Score 17.4; DB 10; Length 495;  
Best Local Similarity 94.7%; Pred. No. 3.9e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agtgtggcactctgtgtg 19  
|||||  
Db 398 AGTGTGGCACCTTTGTGG 380

RESULT 5  
BF711205/c  
LOCUS BF711205.1 496 bp mRNA linear EST 02-JAN-2001  
DEFINITION MI-P-AV1-nrd-c-06-0-UI-s1 MI-P-AV1 Sus scrofa cDNA clone  
KEYWORDS MI-P-AV1-nrd-c-06-0-UI 3', mRNA sequence.  
ACCESSION BF711205.1 GI:12010682  
VERSION BF711205.1  
KEYWORDS EST.  
SOURCE pig.  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
REFERENCE 1 (bases 1 to 496)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE  
JOURNAL  
MEDLINE  
COMMENT

Normalization and subtraction: two approaches to facilitate gene discovery  
Genome Res. 6 (9), 791-806 (1996)  
97044477  
Contact: Tuggle CK  
Molecular Genetics Laboratory, Department of Animal Science  
Iowa State University  
201 Kildee Hall, Ames, IA 50011-3150, USA  
Tel: 5152944252  
Fax: 5152942401  
Email: cktuggle@iastate.edu  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to verify it as a clone from the  
non-normalized placenta library cDNA Library Preparation: M.B. Soares  
Lab, University of Iowa EST sequencing: M.B. Soares Lab, University  
of Iowa Clone distribution: clones will be available through  
Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA=Yes.

FEATURES Location/Qualifiers  
source 1..496  
/organism="Sus scrofa"  
/strain="crossbreed"  
/db\_xref="taxon:9823"  
/clone="MI-P-AV1-nrd-c-06-0-UI"  
/clone\_lib="MI-P-AV1"  
/lab\_host="DH10B (Life Technologies)"  
/note="vector: p7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: EcoRI; The MI-P-AV1  
library is normalized library derived from the MI-P-AV1  
library, ultimately derived from placenta tissue. For a  
detailed description of the library from which this clone  
was derived, please visit our web site at  
http://pigest.genome.iastate.edu/. The procedure used to  
create this library has been previously described (Bonaldo  
Lennon and Soares, Genome Research 6: 791-806, 1996)  
TAG\_LIB=MI-P-AV1  
TAG\_TISSUE=placenta  
TAG\_SEQ=ATTGG"  
BASE COUNT 188 a 86 c 58 g 164 t  
ORIGIN  
Query Match 91.6%; Score 17.4; DB 10; Length 496;  
Best Local Similarity 94.7%; Pred. No. 3.9e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agtgttggaactctgtgtg 19  
|||||  
Db 400 AGTGTGGCACCTTTGTGG 382

RESULT 6  
BF712784/c  
LOCUS BF712784.1 504 bp mRNA linear EST 02-JAN-2001  
DEFINITION MI-P-H1-adu-c-06-1-UM-s1 MI-P-H1 Sus scrofa cDNA clone  
KEYWORDS MI-P-H1-adu-c-06-1-UM 3', mRNA sequence.  
ACCESSION BF712784.1 GI:12012259  
VERSION BF712784.1  
KEYWORDS EST.  
SOURCE pig.  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
REFERENCE 1 (bases 1 to 504)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477

## COMMENT

Contact: Tuggle CK  
Molecular Genetics Laboratory, Department of Animal Science  
Iowa State University  
201 Kildee Hall, Ames, IA 50011-3150, USA  
Tel: 5152944252  
Fax: 5152942401

Email: cktuggle@iastate.edu

The sequence contained an oligo-dt track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dt track served to verify it as a clone from the non-normalized hypothalamus at estrus day 0 library cDNA library preparation: RJ Woods, JA Green, RS Prather S142 Animal Science Research Center, Department of Animal Science, University of Missouri-Columbia, 65211 Clone distribution: clones will be available through Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA=Yes

FEATURES  
source

Location/Qualifiers

1..504  
/organism="Sus scrofa"  
/strain="crossbreed"  
/db\_xref="taxon:9823"  
/clone="MI-P-H1-adu-c-06-1-UM"  
/clone\_lib="MI-P-H1"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: EcoRI; The MI-P-H1 library is derived from hypothalamus at estrus day 0. For a detailed description of the library from which this clone was derived, please visit our web site at <http://pigest.genome.iastate.edu/>.  
TAG\_Lib=MI-P-H1  
TAG\_TISSUE=hypothalamus at estrus day 0  
TAG\_SEQ=TAGATG

BASE COUNT 194 a 61 g 162 t

ORIGIN

## Query Match

Best Local Similarity 91.6%; Score 17.4; DB 10; Length 504;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agtgtggcactctgtg 19

|||||

Db 397 AGTGTGGCATTCTGTGG 379

## RESULT 7

BM432649

LOCUS

DEFINITION BM432649 515 bp mRNA linear EST J1-JAN-2002 LJEJ12C7.ab1 Bos taurus Jejunum #1 library Bos taurus cDNA, mRNA sequence.

ACCESSION BM432649

VERSION BM432649.1 GI:18454371

KEYWORDS EST.

SOURCE COW.

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovidae; Bovinae; Bos.

1 (bases 1 to 515)

Hansen, C., Fu, A., Meng, Y., Li, C., Okine, E., Sensen, C. W., Gordon

, P. M. K. and Moore, S. S.

Gene Expression Profiling of the Bovine Gastrointestinal Tract

Unpublished (2002)

Contact: Dr. Stephen Moore

. Beef Genomics Laboratory

Dept of AFNS, University of Alberta

410 Agri/For, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada

Tel: 780 492 0169

Fax: 780 492 4265

Email: smoores@afns.ualberta.ca  
Insert Length: 515 Std Error: 0.00

POLYA-No.

## FEATURES

source

Location/Qualifiers

1..515  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="Bos taurus Jejunum #1 library"  
/tissue\_type="Smooth muscle"  
/cell\_type="Simple columnar epithelial"  
/dev\_stage="Young adult"  
/lab\_host="XL1-BlueMRF'strain"  
/note="Organ: Intestine; Jejunum; Vector: Uni-22APXR;  
Site\_1: EcoRI; Site\_2: Xho I"

BASE COUNT 180 a 77 c 90 g 168 t

ORIGIN

## Query Match

Best Local Similarity 91.6%; Score 17.4; DB 10; Length 515;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agtgtggcactctgtg 19

|||||

Db 416 ACTGTGGCATTCTGTGG 434

## RESULT 8

BF704178/c

LOCUS

DEFINITION BF704178 522 bp mRNA linear EST 22-DEC-2000 MI-P-03-aaw-d-10-1-UM.sl MI-P-03 Sus scrofa cDNA clone

MI-P-03-aaw-d-10-1-UM 3', mRNA sequence.

ACCESSION BF704178

VERSION BF704178.1 GI:11989586

KEYWORDS EST.

SOURCE pig.

ORGANISM

Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

1 (bases 1 to 522)

Bonaldo, M. F., Lennon, G. and Soares, M. B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

9704477

COMMENT

Contact: Tuggle CK

Molecular Genetics Laboratory, Department of Animal Science

Iowa State University

201 Kildee Hall, Ames, IA 50011-3150, USA

Tel: 5152944252

Fax: 5152942401

Email: cktuggle@iastate.edu

The sequence contained an oligo-dt track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dt track served to verify it as a clone from the

non-normalized ovary at estrus day 12 library cDNA library

preparation: RJ Woods, JA Green, RS Prather S142 Animal Science

Research Center, Department of Animal Science, University of

Missouri-Columbia, 65211 Clone distribution: clones will be

available through Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA=Yes.

## FEATURES

source

Location/Qualifiers

1..522  
/organism="Sus scrofa"  
/strain="crossbreed"  
/db\_xref="taxon:9823"  
/clone="MI-P-03-aaw-d-10-1-UM"  
/clone\_lib="MI-P-03"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-pac (Pharmacia) with a modified

polylinker; Site\_1: Not I; Site\_2: EcoRI; The MI-P-03 library is derived from ovary at estrus day 12. For a detailed description of the library from which this clone was derived, please visit our web site at <http://pigest.genome.iastate.edu/>.

TAG\_LIB=MI-P-03  
TAG\_TISSUE=ovary at estrus day 12  
TAG\_SEQ=TTGTAC\*

BASE COUNT 196 a 88 c 68 g 170 t  
ORIGIN

Query Match 91.6%; Score 17.4; DB 10; Length 522;

Best Local Similarity 94.7%; Pred. No. 3.9e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 agtgtggcactctgtgg 19  
|||||

Db 398 AGTGTGGCACTTTGTGG 380

RESULT 9  
LOCUS BM286313 545 bp mRNA linear EST 28-DEC-2001  
DEFINITION 526314 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BM286313  
VERSION BM286313.1 GI:17995339  
KEYWORDS EST.  
SOURCE COW.

ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 545)  
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J.,ahrenkrug,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G., Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and Keele,J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

JOURNAL Genome Res. 11 (4), 626-630 (2001)  
MEDLINE 21180013  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390

Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -minmatch 12 options.

PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCCGACGACGACG  
Plate: 132 row: 0 Column: 15  
Seq primer: ATTTAGTGACACTATAG.

FEATURES  
source  
1..545  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 3BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle."

BASE COUNT 193 a 82 c 98 g 171 t 1 others

Query Match 91.6%; Score 17.4; DB 10; Length 545;

Best Local Similarity 94.7%; Pred. No. 4e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 agtgtggcactctgtgg 19  
|||||

Db 485 AGTGTGGCACTTTGTGG 503

RESULT 10  
LOCUS BE683254 551 bp mRNA linear EST 25-APR-2001  
DEFINITION 182725 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BE683254  
VERSION BE683254.1 GI:10069938  
KEYWORDS EST.  
SOURCE COW.

ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 551)  
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J.,ahrenkrug,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G., Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and Keele,J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

JOURNAL Genome Res. 11 (4), 626-630 (2001)  
MEDLINE 21180013  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390

Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -minmatch 12 options.

PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCCGACGACGACG  
Plate: 84 row: K Column: 2  
Seq primer: ATTTAGTGACACTATAG.

FEATURES  
source  
1..551  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 4BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from day 20 and day 40 embryos."

BASE COUNT 185 a 82 c 96 g 188 t

ORIGIN

Query Match 91.6%; Score 17.4; DB 10; Length 551;

Best Local Similarity 94.7%; Pred. No. 4e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 agtgtggcactctgtgg 19  
|||||

Db 381 AGTGTGGCACTTTGTGG 399

RESULT 11  
LOCUS BE750214 554 bp mRNA linear EST 25-APR-2001  
DEFINITION 201425 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BE750214  
VERSION BE750214.1 GI:10164206

KEYWORDS EST.  
SOURCE COW.  
ORGANISM Bos taurus

REFERENCE  
AUTHORS Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Perte, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and Keele, J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

JOURNAL MEDLINE 21180013  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt\_trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -minmatch 12 options.

PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCTCAGTCACGACG  
Plate: 111 row: D column: 6  
Seq primer: ATTTAGGTGACACTATAG.  
Location/Qualifiers  
1. 554  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 4BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from day 20 and day 40 embryos."

BASE COUNT 182 a 84 c 97 g 191 t  
ORIGIN

Query Match 91.6%; Score 17.4; DB 10; Length 554;  
Best Local Similarity 94.7%; Pred. No. 4e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agtgttgacactctctgtgg 19  
|||||  
DB 377 AGTGTGGCACTGTGTGG 395

RESULT 12  
LOCUS BE683263  
DEFINITION 182738 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BE683263  
VERSION BE683263.1 GI:10069956  
KEYWORDS EST.  
SOURCE COW.  
ORGANISM Bos taurus

REFERENCE  
AUTHORS Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Perte, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and Keele, J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle

JOURNAL MEDLINE 21180013  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt\_trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -minmatch 12 options.

PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCTCAGTCACGACG  
Plate: 84 row: M column: 3  
Seq primer: ATTTAGGTGACACTATAG.  
Location/Qualifiers  
1. 556  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 4BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from day 20 and day 40 embryos."

BASE COUNT 185 a 83 c 98 g 190 t  
ORIGIN

Query Match 91.6%; Score 17.4; DB 10; Length 556;  
Best Local Similarity 94.7%; Pred. No. 4e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agtgttgacactctctgtgg 19  
|||||  
DB 381 AGTGTGGCACTGTGTGG 399

RESULT 13  
LOCUS BF759242  
DEFINITION MR4-CT0537-051200-005-c07 CT0537 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF759242  
VERSION BF759242.1 GI:12107142  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (Bases 1 to 628)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL MEDLINE 20202663  
COMMENT Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR4st2-MR4-CT0537-051200-005-c07st3-2000-12-05&t4=1)

Seq primer: puc 18 forward  
High quality sequence start: 28  
High quality sequence stop: 619.  
Location/Qualifiers

## FEATURES

source

1. .628  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="C70537"  
/dev\_stage="Adult"

/note="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 236 a 147 c 109 g 135 t 1 others  
ORIGIN

Query Match 91.6%; Score 17.4; DB 10; Length 628;  
Best Local Similarity 94.7%; Pred. No. 4.1e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agtgttgccactctgtg 19

||||| ||||| ||||| ||||| |||||

Db 600 AGGTTCGCATTCTCTGG 618

## RESULT 14

A0942809/c

LOCUS

DEFINITION A0942809 159 bp DNA linear GSS 27-JAN-2000  
Sheared DNA-36123-TR Sheared DNA Trypanosoma brucei genomic clone

ACCESSION A0942809 Sheared DNA-36123, DNA sequence.

VERSION A0942809.1 GI:6766162

KEYWORDS GSS.

SOURCE Trypanosoma brucei.

ORGANISM Trypanosoma brucei

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma

1 (bases 1 to 159)

El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C., Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J., Fraser,C. and Adams,M.

Determination of clone end sequences from Trypanosoma brucei GUTat

10.1 sheared DNA library

Unpublished (1999)

Other\_GSSs: Sheared DNA-36123.TF

Contact: Najib M. El-Sayed

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: nelsayed@tigr.org

Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared

DNA library constructed at TIGR. Clones will be available for

distribution through ATCC. Sheared DNA end sequences search page:

<http://www.tigr.org/tldb/mdb/tbdb/>.

Seq primer: M13-Reverse

Class: shotgun.

## FEATURES

source

Location/Qualifiers

1. .159

/organism="Trypanosoma brucei"

/strain="TREU927/4 GUTat 10.1"

/db\_xref="taxon:5691"

/clone="Sheared DNA-36123"

/clone\_lib="Sheared DNA"

/note="Vector: pUC18; Site\_1: SmaI; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from a cloned population of

Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically

sheared to give a tight size distribution (approx 2 kb).  
The v + i method used for the library construction is  
described in detail in Smith, H.O. and Venter, J.C.  
(Making small insert libraries for whole genome shotgun  
sequencing projects. In Genome Sequencing: A Practical  
Approach, eds. M. Vaudin and B. Borell, Oxford University  
Press, 1999)."

BASE COUNT 46 a 46 c 29 g 38 t  
ORIGIN

Query Match 89.5%; Score 17; DB 12; Length 159;  
Best Local Similarity 100.0%; Pred. No. 4.4e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gtgttgccactctgtg 18

||||| ||||| ||||| ||||| |||||

Db 112 GTGTGGCATTCTGTG 96

## RESULT 15

AA694624

LOCUS

DEFINITION AA694624 567 bp mRNA linear EST 18-DEC-1997  
ET2209 Trypanosoma brucei rhodesiense p II library Trypanosoma

brucei rhodesiense cDNA 5', mRNA sequence.

ACCESSION AA694624

VERSION AA694624.1 GI:2697134

KEYWORDS EST.

SOURCE Trypanosoma brucei rhodesiense.

ORGANISM Trypanosoma brucei rhodesiense

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma

1 (bases 1 to 567)

Ullu,E. and Tschudi,C.

Expressed sequence tags from procyclic Trypanosoma brucei

rhodesiense cDNA clones

Unpublished (1997)

COMMENT

Contact: Ullu E

Department of Internal Medicine, Section of Infectious Diseases

Yale University School of Medicine

P.O. Box 208022, 333 Cedar Street, New Haven, CT 06520-8022, USA

Fax: 203 785 3864

Email: elisabetta.ullu@yale.edu

Seq primer: SK.

Location/Qualifiers

1. .367

/organism="Trypanosoma brucei rhodesiense"

/strain="YTat 1.1"

/db\_xref="taxon:31286"

/clone\_lib="Trypanosoma brucei rhodesiense ZAP II library"

/dev\_stage="Insect form"

/note="Vector: Lambda ZAP II; Site\_1: Eco RI; Site\_2: Xho

I; A unidirectional oligo dt-primer cDNA library was

constructed in lambda ZAP II. Clones were selected using

the criteria of low reactivity with a total cDNA probe.

3 others

BASE COUNT 142 a 111 c 156 g 155 t

ORIGIN

Query Match 89.5%; Score 17; DB 9; Length 567;  
Best Local Similarity 100.0%; Pred. No. 6.2e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gtgttgccactctgtg 18

||||| ||||| ||||| ||||| |||||

Db 346 GTGTGGCATTCTGTG 362

Search completed: May 18, 2002, 17:22:22

Job time: 14693 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:26:35 ; Search time 250.21 Seconds  
(without alignments)  
18.652 Million cell updates/sec

Title: US-09-623-329-24

Sequence: 19  
1 aglgttgacactctgtgg 19

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
C 1	15.8	83.2	4084	2	US-08-568-459A-1
C 2	15.8	83.2	4084	2	US-08-487-826B-1
C 3	15.8	83.2	4084	6	5198347-5
C 4	15.8	83.2	152331	3	US-09-128-155-16
C 5	15.8	83.2	176373	3	US-09-128-155-17
C 6	15.4	81.1	1189	3	US-08-836-582-1
C 7	15.4	81.1	1489	4	US-09-265-566-1
C 8	15.8	77.9	373	1	US-08-451-715A-61
C 9	14.8	77.9	373	2	US-08-967-101-17
C 10	14.8	77.9	373	2	US-08-592-541-17
C 11	14.8	77.9	373	3	US-09-124-698-17
C 12	14.8	77.9	373	4	US-08-127-480-17
C 13	14.8	77.9	373	4	US-08-496-841C-17
C 14	14.8	77.9	920	3	US-09-258-373-2
C 15	14.8	77.9	1875	3	US-09-258-373-21
C 16	14.8	77.9	3164	1	US-08-188-228-49
C 17	14.8	77.9	3164	1	US-08-332-643-43
C 18	14.8	77.9	3164	1	US-08-332-638-49
C 19	14.4	75.8	894	2	US-08-467-963C-28
C 20	14.4	75.8	894	2	US-08-838-189D-28
C 21	14.4	75.8	894	3	US-08-852-344D-28
C 22	14.4	75.8	894	3	US-08-344-639E-28
C 23	14.4	75.8	920	2	US-08-467-963C-7
C 24	14.4	75.8	920	2	US-08-838-189D-7
C 25	14.4	75.8	920	3	US-08-852-344D-7
C 26	14.4	75.8	920	3	US-08-344-639E-7
C 27	14.4	75.8	920	4	US-08-467-969A-7

C 28	14.4	75.8	920	4	US-08-467-961A-7	Sequence 7, Appl
C 29	14.4	75.8	920	4	US-08-001-554A-7	Sequence 7, Appl
C 30	14.4	75.8	946	6	US-08-776-210-1	Sequence 1, Appl
C 31	14.4	75.8	1287	6	5219753-1	Patent No. 519753
C 32	14.4	75.8	2091	6	5190871-3	Patent No. 5190871
C 33	14.4	75.8	3401	6	5190871-1	Patent No. 5190871
C 34	14.4	75.8	15222	4	US-08-801-898A-23	Sequence 23, Appl
C 35	14.4	75.8	15222	4	US-08-962-690-12	Sequence 12, Appl
C 36	14.4	75.8	15223	2	US-08-892-403A-1	Sequence 1, Appl
C 37	14.4	75.8	15223	4	US-08-720-132-1	Sequence 1, Appl
C 38	14.2	74.7	1854	5	PCT-US94-01101-1	Sequence 1, Appl
C 39	14.2	74.7	2263	1	US-08-176-126B-1	Sequence 1, Appl
C 40	14.2	74.7	2263	1	US-08-669-435-1	Sequence 1, Appl
C 41	14.2	74.7	2263	5	PCT-US94-14431A-1	Sequence 1, Appl
C 42	14.2	74.7	3810	2	US-08-475-844-8	Sequence 8, Appl
C 43	14.2	74.7	3810	5	PCT-US95-08429-8	Sequence 8, Appl
C 44	14.2	74.7	4252	2	US-08-475-844-4	Sequence 4, Appl
C 45	14.2	74.7	4252	5	PCT-US95-08429-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1  
US-08-568-459A-1/c  
; Sequence 1, Application US/08568459A  
; Patent No. 5849306  
; GENERAL INFORMATION:  
; APPLICANT: Sim, Kim L.  
; APPLICANT: Chitnois, Chetan  
; APPLICANT: Miller, Louis H.  
; APPLICANT: Peterson, David S.  
; APPLICANT: Su, Xin-zhaun  
; APPLICANT: Wellens, Thomas E.  
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe Martens Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: California  
; COUNTRY: US  
; ZIP: 92660

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICANT NUMBER: US/08/568/459A  
FILING DATE: 07-DEC-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4084 base pairs  
TYPE: nucleic acid  
STRADEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium vivax  
; US-08-568-459A-1



```
Query Match      83.2%; Score 15.8; DB 2; Length 4084;
Best Local Similarity 89.5%; Pred. No. 25;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 agtgttggcacttctgtgg 19
    ||| || ||||| |||||
Db 1593 AGTTTGGCACTTCTGTGG 1575

RESULT 2
US-08-487-826B-1/c
; Sequence 1, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release
; CURRENT APPLICATION NUMBER: US/08/487.826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4084 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
US-08-487-826B-1

Query Match      83.2%; Score 15.8; DB 2; Length 4084;
Best Local Similarity 89.5%; Pred. No. 25;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 agtgttggcacttctgtgg 19
    ||| || ||||| |||||
Db 1593 AGTTTGGCACTTCTGTGG 1575

RESULT 3
5198347-5/c
; Patent No. 5198347
; APPLICANT: Miller, LOUIS H.; ADAMS, JOHN H.; KASLOW,
; DAVIC C.; FANG, XIANGDOUG
```

```
; TITLE OF INVENTION: DNA ENCODING PLASMODIUM VIVAX AND
; PLASMODIUM KNOWLES DUFFY RECEPTOR
; NUMBER OF SEQUENCES: 27
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/554,837
; FILING DATE: 20-JUL-1990
; SEQ ID NO: 5
; LENGTH: 4084
5198347-5

Query Match      83.2%; Score 15.8; DB 6; Length 4084;
Best Local Similarity 89.5%; Pred. No. 25;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 agtgttggcacttctgtgg 19
    ||| || ||||| |||||
Db 1593 AGTTTGGCACTTCTGTGG 1575

RESULT 4
US-09-128-155-16/c
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match      83.2%; Score 15.8; DB 3; Length 152331;
Best Local Similarity 89.5%; Pred. No. 49;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 agtgttggcacttctgtgg 19
    ||| || ||||| |||||
Db 137565 ACTGTAGGCACCTTCTGTGG 137547

RESULT 5
US-09-128-155-17/c
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; NUMBER OF SEQ ID NOS: 18
```

; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 17  
; LENGTH: 176373  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(176373)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-128-155-17

Query Match 83.2%; Score 15.8; DB 3; Length 176373;  
Best Local Similarity 89.5%; Pred. No. 50;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 agtgtaggcacttctgtg 19  
|||||  
DB 150146 ACTGTAGGCACCTTCTGTG 150128

RESULT 6  
US-08-836-582-1  
; Sequence 1, Application US/08836582  
; Patent No. 6045999  
; GENERAL INFORMATION:  
; APPLICANT: Bernards, Rene  
; APPLICANT: Beijersbergen, Roderick L  
; TITLE OF INVENTION: Transcription factor E2F-4  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon & Vanderhye  
; STREET: 1100 No. 6045999th Glebe Road, 8th Floor  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: US  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/836,582  
; FILING DATE: 14-MAY-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB95/00868  
; FILING DATE: 18-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9423049.7  
; FILING DATE: 15-NOV-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arthur R. Crawford  
; REGISTRATION NUMBER: 25,327  
; REFERENCE/DOCKET NUMBER: 620-18  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1489 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 27..1268  
US-08-836-582-1

Query Match 81.1%; Score 15.4; DB 3; Length 1489;  
Best Local Similarity 94.1%; Pred. No. 34;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 gtgtggcacttctgtg 18

DB 1429 GTGCTGGCAGCTTCTGTG 1445  
|||||

RESULT 7  
US-09-265-566-1  
; Sequence 1, Application US/09265566  
; Patent No. 6303335  
; GENERAL INFORMATION:  
; APPLICANT: Bernards, Rene  
; APPLICANT: Beijersbergen, Roderick L  
; TITLE OF INVENTION: Transcription factor E2F-4  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon & Vanderhye  
; STREET: 1100 No. 6303335th Glebe Road, 8th Floor  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: US  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/265,566  
; FILING DATE: 10-Mar-1999  
; CLASSIFICATION: <Unknown>  
; 15-NOV-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/836,582  
; FILING DATE: <Unknown>  
; APPLICATION NUMBER: GB 9423049.7  
; FILING DATE: 15-NOV-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arthur R. Crawford  
; REGISTRATION NUMBER: 25,327  
; REFERENCE/DOCKET NUMBER: 620-18  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1489 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 27..1268  
; SEQUENCE DESCRIPTION: Seq ID No: 1:  
US-09-265-566-1

Query Match 81.1%; Score 15.4; DB 4; Length 1489;  
Best Local Similarity 94.1%; Pred. No. 34;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gtgtggcacttctgtg 18  
|||||

DB 1429 GTGCTGGCAGCTTCTGTG 1445  
|||||

RESULT 8  
US-08-451-715A-61/c  
; Sequence 61, Application US/08451715A  
; Patent No. 5801013  
; GENERAL INFORMATION:  
; APPLICANT: Tao, Jianshi  
; APPLICANT: Qiu, Yan  
; APPLICANT: Houman, Fariba  
; APPLICANT: Shen, Xiaoyu  
; APPLICANT: Schimmel, Paul R.  
; TITLE OF INVENTION: Helicobacter Aminoacyl-tRNA Synthetase  
; TITLE OF INVENTION: Proteins, Nucleic Acids and Strains Comprising Same

NUMBER OF SEQUENCES: 67  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/451.715A  
FILING DATE: 26-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BROOK, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: C9194-25  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 61:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 33 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-451-715A-61

Query Match 78.9%; Score 15; DB 1; Length 33;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 gttggcacttctgtg 18  
Db 29 GTTGGCACTTCTGTG 15

RESULT 9  
US-08-967-101-17  
Sequence 17, Application US/08967101  
Patent No. 5840540  
GENERAL INFORMATION:  
APPLICANT: ST. GEORGE-HYSLOP, PETER H  
APPLICANT: ROMMENS, JOHANNA M  
APPLICANT: FRASER, PAUL E  
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
TO ALZHEIMER'S DISEASE  
NUMBER OF SEQUENCES: 183  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
STREET: High Street Tower - 125 High Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/967.101  
FILING DATE: 10-NOV-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/592,541  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:

NAME: Pitcher, Edmund R.  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7000  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 373 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-967-101-17

Query Match 77.9%; Score 14.8; DB 2; Length 373;  
Best Local Similarity 84.2%; Pred. No. 54;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 agtgtggcacttctgtgg 19  
Db 310 ANTCTGGCACTTCTATGG 328

RESULT 10  
US-08-592-541-17  
Sequence 17, Application US/08592541  
Patent No. 5986054  
GENERAL INFORMATION:  
APPLICANT: ST. GEORGE-HYSLOP, PETER H  
APPLICANT: ROMMENS, JOHANNA M  
APPLICANT: FRASER, PAUL E  
TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED  
TO ALZHEIMER'S DISEASE  
NUMBER OF SEQUENCES: 183  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TESTA, HURWITZ & THIBEAULT  
STREET: High Street Tower - 125 High Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/592,541  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Pitcher, Edmund R.  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7000  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 373 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-592-541-17

Query Match 77.9%; Score 14.8; DB 2; Length 373;  
Best Local Similarity 84.2%; Pred. No. 54;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 agtgtggcacttctgtgg 19  
Db 310 ANTCTGGCACTTCTATGG 328

```
RESULT 11
US-09-124-698-17
; Sequence 17, Application US/09124698
; Patent No. 6117978
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/124,698
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/592,541
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 373 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-124-698-17

Query Match 77.9%; Score 14.8; DB 3; Length 373;
Best Local Similarity 84.2%; Pred. No. 54;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 agtggtagcactctctg 19
| | | | | | | | | | |
Db 310 ANTCTGGCAGCTTCTATGG 328

RESULT 12
US-09-127-480-17
; Sequence 17, Application US/09127480
; Patent No. 6194153
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 183
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/496,841C
; FILING DATE: 28-Jun-1995
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul F. Fehlner, Ph.D.
; REGISTRATION NUMBER: 35,135
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 527-7700
; TELEFAX: (212) 753-6237
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 373 base pairs
; TYPE: nucleic acid
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/127,480
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,541
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 373 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-127-480-17

Query Match 77.9%; Score 14.8; DB 4; Length 373;
Best Local Similarity 84.2%; Pred. No. 54;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 agtggtagcactctctg 19
| | | | | | | | | | |
Db 310 ANTCTGGCAGCTTCTATGG 328

RESULT 13
US-08-496-841C-17
; Sequence 17, Application US/08496841C
; Patent No. 6210919
; GENERAL INFORMATION:
; APPLICANT: ST. GEORGE-HYSLOP, PETER H
; APPLICANT: ROMMENS, JOHANNA M
; APPLICANT: FRASER, PAUL E
; TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
; TO ALZHEIMER'S DISEASE
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby, PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/496,841C
; FILING DATE: 28-Jun-1995
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul F. Fehlner, Ph.D.
; REGISTRATION NUMBER: 35,135
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 527-7700
; TELEFAX: (212) 753-6237
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 373 base pairs
; TYPE: nucleic acid
```

Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 gtgttgccacttctgtgg 19  
| | | | | | | | | | | | | | | | | |  
Db 358 GTGTGGCAGCTTCGCTGG 341

Search completed: May 18, 2002, 17:26:41  
Job time: 12843 sec

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 17:  
US-08-496-841C-17

Query Match 77.9%; Score 14.8; DB 4; Length 373;  
Best Local Similarity 84.2%; Pred. No. 54;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 agtgttgccacttctgtgg 19  
| | | | | | | | | | | | | | | | | |  
Db 310 ANTCTGGCAGCTTCATGG 328

RESULT 14  
US-09-258-373-2/c  
; Sequence 2, Application US/09258373  
; Patent No. 6150110  
; GENERAL INFORMATION:  
; APPLICANT: Fletcher, Jonathan A.  
; APPLICANT: Xiao, Sheng  
; TITLE OF INVENTION: HMGI(Y)-LAMA4\* FUSION ONCOGENE,  
; FILE REFERENCE: B0801/7135/ERP  
; CURRENT APPLICATION NUMBER: US/09/258,373  
; EARLIER FILING DATE: 1999-02-26  
; EARLIER APPLICATION NUMBER: 60/076,401  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 920  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
US-09-258-373-2

Query Match 77.9%; Score 14.8; DB 3; Length 920;  
Best Local Similarity 88.9%; Pred. No. 63;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 gtgttgccacttctgtgg 19  
| | | | | | | | | | | | | | | | | |  
Db 358 GTGTGGCAGCTTCGCTGG 341

RESULT 15  
US-09-258-373-21/c  
; Sequence 21, Application US/09258373  
; Patent No. 6150110  
; GENERAL INFORMATION:  
; APPLICANT: Fletcher, Jonathan A.  
; APPLICANT: Xiao, Sheng  
; TITLE OF INVENTION: HMGI(Y)-LAMA4\* FUSION ONCOGENE,  
; FILE REFERENCE: B0801/7135/ERP  
; CURRENT APPLICATION NUMBER: US/09/258,373  
; EARLIER FILING DATE: 1999-02-26  
; EARLIER APPLICATION NUMBER: 60/076,401  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 21  
; LENGTH: 1875  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
US-09-258-373-21

Query Match 77.9%; Score 14.8; DB 3; Length 1875;  
Best Local Similarity 88.9%; Pred. No. 72;

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:45:58 ; Search time 1024.22 Seconds  
(without alignments)  
31.850 Million cell updates/sec

Title: US-09-623-329-24

Perfect score: 19

Sequence: 1 agtgttgacactctgtgg 19

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

- 1: /SID55/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SID55/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SID55/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SID55/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SID55/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SID55/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SID55/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SID55/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SID55/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SID55/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SID55/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SID55/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SID55/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SID55/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SID55/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SID55/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	19	100.0	19	20	AAZ11674
C	16.4	86.3	355	21	AA44026
3	16.4	86.3	4281	23	ABL22852
C	15.8	83.2	264	21	AAC22179
5	15.8	83.2	952	21	AAC45824
6	15.8	83.2	955	21	AAC34136
C	15.8	83.2	2230	23	ABL11228
8	15.8	83.2	2687	22	AAS36599
9	15.8	83.2	2687	22	AAS36600

C	10	15.8	83.2	2957	23	ABL14983
C	11	15.8	83.2	3114	22	AA01375
C	12	15.8	83.2	3298	17	AAAT45082
C	13	15.8	83.2	4084	12	AA013317
C	14	15.8	83.2	4084	16	AA083524
C	15	15.8	83.2	4084	18	AAAT72889
C	16	15.8	83.2	4084	21	AAZ98281
C	17	15.8	83.2	4309	23	AAZ79116
C	18	15.8	83.2	5026	23	ABL14982
C	19	15.8	83.2	5438	22	AAH47054
C	20	15.8	83.2	6101	22	AAH47055
C	21	15.8	83.2	7885	22	AAAT70210
C	22	15.8	83.2	8760	15	AAQ73473
C	23	15.8	83.2	8855	20	AAAO2997
C	24	15.8	83.2	28866	20	AAAZ2304
C	25	15.4	81.1	439	14	AAQ60111
C	26	15.4	81.1	570	23	AAZ91577
C	27	15.4	81.1	819	21	AACT9861
C	28	15.4	81.1	866	21	AAAF3214
C	29	15.4	81.1	1489	17	AAAT29618
C	30	15.4	81.1	1489	19	AAV17828
C	31	15.4	81.1	1539	23	AAZ91970
C	32	15.4	81.1	2222	24	AAZ62261
C	33	15.4	81.1	39567	22	AAK74053
C	34	15	78.9	33	19	AAV53188
C	35	15	78.9	7868	22	AAZ42112
C	36	14.8	77.9	375	16	AAAT24623
C	37	14.8	77.9	387	21	AAAO2017
C	38	14.8	77.9	465	22	AAI93321
C	39	14.8	77.9	513	21	AAAC41198
C	40	14.8	77.9	610	21	AAAC98357
C	41	14.8	77.9	672	23	AAZ81287
C	42	14.8	77.9	920	22	AAAC31803
C	43	14.8	77.9	1179	21	AAAC34895
C	44	14.8	77.9	1227	21	AAAC52020
C	45	14.8	77.9	1317	22	AAAO5135

#### ALIGNMENTS

#### RESULT 1

AAZ11674

ID AAZ11674 standard; DNA; 19 BP.

AC AAZ11674;

DT 19-NOV-1999 (first entry)

DE Oligo specific for EBV BARF-1 RNA.

KW Epstein Barr Virus; EBV infection; viral; gene transcription; EBER-1;  
KW Epstein Barr early RNA; Epstein Barr nuclear antigen 1; EBNA-1; LMP-1;  
KW latent membrane protein; LMP-2; vIL10; BCRF-1; BARF1; BDLF2; NASBA;  
KW EBV-associated malignancy; primer; ss.

OS Synthetic.

OS Epstein-barr virus.

PN WO9945155-A2.

PD 10-SEP-1999.

PF 01-MAR-1999; 99WO-EP01392.

PR 04-MAR-1998; 98EP-0200655.

PR 14-DEC-1998; 98EP-0204231.

PA (ALKU ) AKZO NOBEL NV.

PI Vervoort MBHJ, Van Den Brule AJC, Middelorp JM;

DR WPI; 1999-551051/46.

```

XX PT Identifying Epstein Barr Virus infection
XX PS Claim 11; Page 20; 50pp; English.
XX CC The invention provides methods for identifying an Epstein Barr Virus
XX (EBV) infection, that comprises determining viral gene transcription
XX patterns by amplification of specific RNA sequences. The binding sites
XX of the oligos suitable for amplification are located in the following
XX genes: Epstein Barr early RNA (EBER-1), Epstein Barr nuclear antigen 1
XX (EBNA-1), latent membrane protein 1 (LMP-1), LMP-2, and villo (BCRF-1),
XX BAF1 and BDLF2. The method comprises (a) amplifying a target sequence
XX within one or more RNA(s) transcribed from above gene sequences and the
XX (b) detecting the amplified products, determining the transcription
XX pattern and identifying the corresponding EBV-associated malignancy. The
XX RNA is amplified using a transcription based amplification technique
XX such as NASBA. The invention is used to diagnose malignant and
XX CC non-malignant EBV-associated diseases. Sequences AA211672-75 represent
XX CC oligos specific for BARF-1 RNA.
XX SQ Sequence 19 BP; 2 A; 3 C; 7 G; 7 T; 0 other;

Query Match 100.0%; Score 19; DB 20; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 agtgtggcactctgtgg 19
DB 1 agtgtggcactctgtgg 19

RESULT 2
AAA44026/C
ID AAA44026 standard; cDNA; 355 BP.
XX AC AAA44026;
XX DT 21-AUG-2000 (first entry)
XX DE Mouse secreted expressed sequence tag SEQ ID NO:601.
XX KW Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;
XX expressed sequence tag; EST; probe; chemotactic; proliferative;
XX immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
XX thrombolytic; antiinflammatory; cytosolic; antibacterial; antifungal;
XX antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;
XX antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
XX cerebroprotective; anticonvulsant; antidepressant; gene therapy;
XX vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
XX insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
XX lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
XX central nervous system disorder; Alzheimer's disease; stroke;
XX Parkinson's disease; Huntington's disease; coagulation disorder;
XX haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
XX tumour; infection; depression; psoriasis; ss.
XX OS Mus musculus.
XX PN WO200021991-A1.
XX PD 20-APR-2000.
XX PF 15-OCT-1999; 99WO-US24206.
XX PR 15-OCT-1998; 98US-0104436.
XX PA (GEMY ) GENETICS INST INC.
XX PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX PI Merberg D, Treacy M, Bowman MR;
XX DR WPI; 2000-317938/27.

Isolated polynucleotides, and encoded proteins, comprising secreted
expressed sequence tags (SESTs), useful for treating various disorders
such as autoimmune, infectious, and central nervous system disorders -
Claim 1; Page 360; 803pp; English.
AAA43426 to AAA45925 represent specifically claimed secreted expressed
sequence tags (SESTs), isolated from human, mouse, chicken and rat
tissue sources. The SESTs can have a range of activities depending on
the tissues they were isolated from. The activities include:
chemotactic; proliferative; immunomodulatory; haematopoietic;
chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
antiasthmatic; vulnary; antiulcer; osteopathic; neuroprotective;
nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
anticonvulsant; and antidepressant. The SESTs can be used for gene
therapy and in vaccines. The SESTs are useful as probes for the
identification and isolation of full-length cDNAs and genomic DNA
molecules which correspond to the SESTs. Proteins encoded by the SESTs
are useful in assays for determining biological activity and raising
antibodies. They may be useful for treatment of autoimmune disorders
(multiple sclerosis, insulin dependent diabetes), allergic conditions
(asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
osteoporosis, osteoarthritis, central nervous system disorders
(Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
disease), tumours, bacterial, fungal or viral infections, depression and
psoriasis. AAA45926 to AAA45931 represent linker variants which are given
in the exemplification of the present invention.
Sequence 355 BP; 89 A; 108 C; 75 G; 83 T; 0 other;

Query Match 86.3%; Score 16.4; DB 21; Length 355;
Best Local Similarity 94.4%; Pred. No. 62;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agtgtggcactctgtg 18
DB 300 AGTGTGGCACCCTGTGTG 283

RESULT 3
ABU22852
ID ABL22852 standard; DNA; 4281 BP.
XX AC ABL22852;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 20029.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.

```

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PS interactions -  
XX  
XX Claim 1; SEQ ID NO 20029; 2lpp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABU16176-ABU16175), expressed DNA  
CC sequences (ABU16176-ABU16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 4281 BP; 1178 A; 1023 C; 1029 G; 1051 T; 0 other;

Query Match 86.3%; Score 16.4; DB 23; Length 4281;  
Best Local Similarity 94.4%; Pred. No. 87;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 2 gtgtggcactctctgtg 19  
|||||  
Db 3680 gtgtggcactctctgtg 3697

RESULT 4  
AAC22179/c  
ID AAC22179 standard; cDNA; 264 BP.  
XX  
XX AAC22179;  
XX  
XX 06-OCT-2000 (first entry)  
XX  
XX Human secreted protein 5' EST, SEQ ID NO: 26254.  
XX  
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX gene therapy; chromosome mapping; ss.  
XX Homo sapiens.  
XX  
XX EP1033401-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 21-FEB-2000; 2000EP-0200610.  
XX  
XX 26-FEB-1999; 99US-0122487.  
XX  
XX (GEST ) GENSET.  
XX  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
XX  
XX WPI; 2000-500381/45.  
XX  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -

PS Claim 1; SEQ ID 26254; 7lpp + CD-ROM; English.  
XX  
XX The present sequence is one of a large number of 5' ESTs derived from  
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively  
XX identified within the present sequence. The 5' ESTs were prepared from  
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
XX sequences usually correspond mainly to the 3' untranslated region (UTR)  
XX of the mRNA because they are often obtained from oligo-dT primed cDNA  
XX libraries. Such ESTs are not well suited for isolating cDNA sequences  
XX derived from the 5' ends of mRNAs and even in those cases where longer

CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.  
XX  
XX Sequence 264 BP; 76 A; 55 C; 58 G; 75 T; 0 other;

Query Match 83.2%; Score 15.8; DB 21; Length 264;  
Best Local Similarity 89.5%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 agtqttagcactctctgtg 19  
|||||  
Db 264 AGTGTTCGACAGTCTGTGG 246

RESULT 5  
AAC45824  
ID AAC45824 standard; DNA; 952 BP.  
XX  
XX AAC45824;  
XX  
XX 18-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 47895.  
XX  
XX Hybridisation assay; genetic mapping; gene expression control;  
XX protein identification; signal transduction pathway;  
XX metabolic pathway; promoter; termination sequence; ss.  
XX Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
XX 09-MAR-1999; 99US-0123548.  
XX 23-MAR-1999; 99US-0125788.  
XX 25-MAR-1999; 99US-0126264.  
XX 29-MAR-1999; 99US-0126785.  
XX 01-APR-1999; 99US-0127462.  
XX 06-APR-1999; 99US-0128234.  
XX 08-APR-1999; 99US-0128714.  
XX 16-APR-1999; 99US-0129845.  
XX 19-APR-1999; 99US-0130077.  
XX 21-APR-1999; 99US-0130449.  
XX 23-APR-1999; 99US-0130510.  
XX 28-APR-1999; 99US-0130891.  
XX 30-APR-1999; 99US-0131449.  
XX 30-APR-1999; 99US-0132048.  
XX 04-MAY-1999; 99US-0132407.  
XX 05-MAY-1999; 99US-0132484.  
XX 06-MAY-1999; 99US-0132485.  
XX 06-MAY-1999; 99US-0132486.  
XX 07-MAY-1999; 99US-0132487.  
XX 11-MAY-1999; 99US-0132863.  
XX 14-MAY-1999; 99US-0134256.  
XX 14-MAY-1999; 99US-0134218.  
XX 14-MAY-1999; 99US-0134219.  
XX 14-MAY-1999; 99US-0134221.  
XX 18-MAY-1999; 99US-0134370.  
XX 19-MAY-1999; 99US-0134768.  
XX 20-MAY-1999; 99US-0134941.  
XX 21-MAY-1999; 99US-0135124.  
XX 24-MAY-1999; 99US-0135353.  
XX 24-MAY-1999; 99US-0135629.



PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.

```
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      83.2%; Score 15.8; DB 21; Length 952;
Best Local Similarity 89.5%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 agtgttgccactctctg 19
    ||||| ||||| |||
Db 156 agtgttgccactctctg 174

RESULT 6
ID AAC34136 standard; DNA; 955 BP.
XX
AC AAC34136;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 5577.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.

10-JUN-1999; 99US-0138094.
10-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847.
14-JUN-1999; 99US-0139119.
16-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
17-JUN-1999; 99US-0139492.
18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.
18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139763.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139899.
23-JUN-1999; 99US-0140353.
23-JUN-1999; 99US-0140354.
24-JUN-1999; 99US-0140695.
28-JUN-1999; 99US-0140823.
29-JUN-1999; 99US-0140991.
30-JUN-1999; 99US-0141287.
01-JUL-1999; 99US-0141842.
01-JUL-1999; 99US-0142154.
02-JUL-1999; 99US-0142055.
06-JUL-1999; 99US-0142390.
08-JUL-1999; 99US-0142803.
09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
16-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.
20-JUL-1999; 99US-0144352.
20-JUL-1999; 99US-0144632.
20-JUL-1999; 99US-0144884.
21-JUL-1999; 99US-0144814.
21-JUL-1999; 99US-0145086.
21-JUL-1999; 99US-0145088.
22-JUL-1999; 99US-0145085.
22-JUL-1999; 99US-0145087.
22-JUL-1999; 99US-0145089.
22-JUL-1999; 99US-0145192.
23-JUL-1999; 99US-0145145.
23-JUL-1999; 99US-0145218.
23-JUL-1999; 99US-0145224.
26-JUL-1999; 99US-0145276.
27-JUL-1999; 99US-0145913.
27-JUL-1999; 99US-0145918.
27-JUL-1999; 99US-0145919.
28-JUL-1999; 99US-0145951.
02-AUG-1999; 99US-0146386.
02-AUG-1999; 99US-0146388.
02-AUG-1999; 99US-0146389.
03-AUG-1999; 99US-0147038.
04-AUG-1999; 99US-0147204.
04-AUG-1999; 99US-0147302.
05-AUG-1999; 99US-0147192.
05-AUG-1999; 99US-0147260.
06-AUG-1999; 99US-0147303.
06-AUG-1999; 99US-0147416.
```

PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159337.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 18-OCT-1999; 99US-0159638.  
 PR 21-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 25-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 83.2%; Score 15.8; DB 21; Length 955;  
 Best Local Similarity 89.5%; Pred. No. 1.4e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 agtgttggcactctgtg 19  
 Db 157 agtgttggcactctg 175  
 RESULT 7  
 ABL11228/c  
 ID ABL11228 standard; cDNA; 2230 BP.  
 XX  
 AC ABL11228;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 28166.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers BW;  
 XX  
 DR WPI: 2001-656860/75.  
 DR P-PSDB; ABB67125.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.  
 XX  
 PS Claim 1; SEQ ID NO 28166; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 2230 BP; 616 A; 461 C; 489 G; 564 T; 0 other;

Query Match 83.2%; Score 15.8; DB 23; Length 2230;  
 Best Local Similarity 89.5%; Pred. No. 1.6e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 agtgttggcactctgtg 19  
 Db 215 AGTGTGGCCTCTCTAAGG 197

RESULT 8  
 AAS36599  
 ID AAS36599 standard; DNA; 2687 BP.  
 XX  
 AC AAS36599;  
 XX  
 DT 17-DEC-2001 (first entry)

XX Human cardiovascular system antigen genomic DNA SEQ ID No 2099.  
DE Cardiovascular antigen; human; mouse; rabbit; goat; horse; cat;  
XX Chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;  
KW Antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;  
KW Cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;  
KW Ophthalmological; vulnary; gene therapy; autoimmune disease; neoplasm;  
KW Hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;  
KW Cerebrovascular disorder; nervous system disorder; bacterial infection;  
KW Fungal infection; viral infection; ocular disorder; endocrine disorder;  
KW Gastrointestinal disorder; renal disorder; respiratory disorder;  
KW Wound healing; skin aging; organ transplantation; tissue regeneration;  
XX Anti-infertility.  
XX  
OS Homo sapiens.  
PN WO200155321-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01340.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205151.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226682.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228024.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 06-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239933.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.



```

Query Match      83.2%;   Score 15.8; DB 22; Length 2687;
Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      ' 1 agtgttggcactctctgtgg 19
          | |||||
Db       2055 attcttggcactctctgtcg 2073

RESULT 10
ABL14983/C
ID ABL14983 standard; cDNA; 2957 BP.
XX
XX
XX
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 39431.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
XX WO200171042-A2.
PN
XX
PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
```

XX 10-SEP-1999; 99US-0393996.  
PR (MILL-) MILLENNIUM PHARM INC.  
XX  
XX Fraser CC, Sharp JD, Wrighton N, Myers PS, Goodearl ADJ;  
PI  
XX WPI: 2001-183280/18.  
DR P-PSDB; AAU00479.  
XX  
XX Isolated nucleic acid molecules encoding proteins useful as modulating  
PT agents in regulating a variety of cellular processes are used for  
PT treating e.g. cancer and autoimmune disorders -  
XX  
XX Claim 2; Fig 6A-6C; 326pp; English.  
XX  
XX The present sequence encoding for human TANGO 405 protein is isolated  
CC from cDNA clone jthLal52h06 from a human mixed lymphocyte reaction cDNA  
CC library. It is 1 of 6 novel human proteins which include TANGO 210  
CC (AAU00469), TANGO 364 (AAU00471), TANGO 366 (AAU00472), INTERCEPT 394  
CC (AAU00473) and INTERCEPT 400 (AAU00476). Novel sequences for murine  
CC TANGO 210 (AAU00470), INTERCEPT 400 (AAU00477), TANGO 405 (AAU00480) and  
CC a rat INTERCEPT 400 (AAU00478) sequence are also described. The nucleic  
CC acids encoding these novel proteins are useful as modulating agents in  
CC regulating a variety of cellular processes and can be used to express  
CC the proteins in a host cell in gene therapy applications. Human and  
CC murine TANGO 405 proteins show sequence homology to murine dectin-2.  
CC TANGO 405 modulates growth, proliferation, survival, differentiation,  
CC activity, morphology and movement/migration of human lymphocytes and  
CC bone marrow cells and tissues and can be used to prevent, diagnose or  
CC treat leukaemia, lymphomas and autoimmune disorders.  
XX  
XX Sequence 3114 BP; 1001 A; 527 C; 517 G; 1069 T; 0 other;  
SQ

Query Match 83.2%; Score 15.8; DB 22; Length 3114;  
Best Local Similarity 89.5%; Pred. No. 1.7e+02; Gaps 0;  
Matches 17; Conservative 0; Mismatches 2; Indels 0

QY 1 agtgttgccactctgtgg 19  
|||||  
DB 2895 AGTGTGGCACCCTTCTGG 2877

RESULT 12  
AAT45082  
ID AAT45082 standard; cDNA to mRNA; 3298 BP.  
XX  
XX AAT45082;  
XX  
XX DT 05-FEB-1997 (first entry)  
XX  
XX Beta-1-4-galactosyltransferase-related protein coding sequence #2.  
DE  
XX Murine; beta-1,4-galactosyltransferase-related protein; sterility;  
KW fertilisation; F9 cancer cell; Huynh's method; ds.  
XX  
XX OS Mus musculus.  
XX  
XX Key Location/Qualifiers  
FH 83..1081  
FT CDS  
FT /\*tag= a  
FT /product= beta-1-4-galactosyltransferase-related protein  
XX JP08196279-A.  
XX  
XX PD 06-AUG-1996.  
XX  
XX PF 25-JAN-1995; 95JP-0009642.  
XX  
XX PR 25-JAN-1995; 95JP-0009642.  
XX  
XX PA (MITK ) MITSUI TOATSU CHEM INC.

XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEXE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI: 2001-656860/75.  
DR P-PSDB; ABB70880.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Claim 1; SEQ ID NO 39431; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 2957 BP; 734 A; 839 C; 770 G; 614 T; 0 other;  
SQ

Query Match 83.2%; Score 15.8; DB 23; Length 2957;  
Best Local Similarity 89.5%; Pred. No. 1.7e+02; Gaps 0;  
Matches 17; Conservative 0; Mismatches 2; Indels 0

QY 1 agtgttgccactctgtgg 19  
|||||  
DB 2353 AGTGTGGCCTCTCTGG 2335

RESULT 11  
AAS01375/C  
ID AAS01375 standard; cDNA; 3114 BP.  
XX  
XX AAS01375;  
XX  
XX DT 04-JUL-2001 (first entry)  
XX  
XX DE Human TANGO 405 cDNA sequence.  
XX  
XX Human; TANGO 210; clone jthLal52h06; TANGO 364; TANGO 366; dectin-2;  
KW INTERCEPT 394; INTERCEPT 400; TANGO 405; cellular process regulator;  
KW gene therapy; growth modulator; proliferation; cell differentiation;  
KW lymphocyte; bone marrow cell migration; leukaemia; lymphoma;  
KW autoimmune disorder; ss.  
XX  
XX OS Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH 154..783  
FT CDS  
FT /\*tag= a  
FT /product= "TANGO 405 protein"  
FT /note= "The ORF is specifically claimed"  
FT sig\_peptide 154..297  
FT mat\_peptide 298..780  
FT /\*tag= b  
FT /\*tag= c  
XX  
XX WO200118016-A1.  
XX  
XX 15-MAR-2001.  
XX  
XX 30-JUN-2000; 2000WO-US18174.  
XX  
XX PF

XX 10-SEP-1999; 99US-0393996.  
PR (MILL-) MILLENNIUM PHARM INC.  
XX  
XX Fraser CC, Sharp JD, Wrighton N, Myers PS, Goodearl ADJ;  
PI  
XX WPI: 2001-183280/18.  
DR P-PSDB; AAU00479.  
XX  
XX Isolated nucleic acid molecules encoding proteins useful as modulating  
PT agents in regulating a variety of cellular processes are used for  
PT treating e.g. cancer and autoimmune disorders -  
XX  
XX Claim 2; Fig 6A-6C; 326pp; English.  
XX  
XX The present sequence encoding for human TANGO 405 protein is isolated  
CC from cDNA clone jthLal52h06 from a human mixed lymphocyte reaction cDNA  
CC library. It is 1 of 6 novel human proteins which include TANGO 210  
CC (AAU00469), TANGO 364 (AAU00471), TANGO 366 (AAU00472), INTERCEPT 394  
CC (AAU00473) and INTERCEPT 400 (AAU00476). Novel sequences for murine  
CC TANGO 210 (AAU00470), INTERCEPT 400 (AAU00477), TANGO 405 (AAU00480) and  
CC a rat INTERCEPT 400 (AAU00478) sequence are also described. The nucleic  
CC acids encoding these novel proteins are useful as modulating agents in  
CC regulating a variety of cellular processes and can be used to express  
CC the proteins in a host cell in gene therapy applications. Human and  
CC murine TANGO 405 proteins show sequence homology to murine dectin-2.  
CC TANGO 405 modulates growth, proliferation, survival, differentiation,  
CC activity, morphology and movement/migration of human lymphocytes and  
CC bone marrow cells and tissues and can be used to prevent, diagnose or  
CC treat leukaemia, lymphomas and autoimmune disorders.  
XX  
XX Sequence 3114 BP; 1001 A; 527 C; 517 G; 1069 T; 0 other;  
SQ

Query Match 83.2%; Score 15.8; DB 22; Length 3114;  
Best Local Similarity 89.5%; Pred. No. 1.7e+02; Gaps 0;  
Matches 17; Conservative 0; Mismatches 2; Indels 0

QY 1 agtgttgccactctgtgg 19  
||||||| |  
DB 2895 AGTGTGGCACCCTTCTGG 2877

RESULT 12  
AAT45082  
ID AAT45082 standard; cDNA to mRNA; 3298 BP.  
XX  
XX AAT45082;  
XX  
XX DT 05-FEB-1997 (first entry)  
XX  
XX Beta-1-4-galactosyltransferase-related protein coding sequence #2.  
DE  
XX Murine; beta-1,4-galactosyltransferase-related protein; sterility;  
KW fertilisation; F9 cancer cell; Huynh's method; ds.  
XX  
XX Mus musculus.  
OS  
XX  
XX Key Location/Qualifiers  
FH 83..1081  
FT CDS  
FT /tag= a  
FT /product= beta-1-4-galactosyltransferase-related protein  
XX JP08196279-A.  
XX  
XX 06-AUG-1996.  
XX  
XX 25-JAN-1995; 95JP-0009642.  
XX  
XX 25-JAN-1995; 95JP-0009642.  
XX  
XX (MITK ) MITSUI TOATSU CHEM INC.  
PA

XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEXE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI: 2001-656860/75.  
DR P-PSDB; ABB70880.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Claim 1; SEQ ID NO 39431; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 2957 BP; 734 A; 839 C; 770 G; 614 T; 0 other;  
SQ

Query Match 83.2%; Score 15.8; DB 23; Length 2957;  
Best Local Similarity 89.5%; Pred. No. 1.7e+02; Gaps 0;  
Matches 17; Conservative 0; Mismatches 2; Indels 0

QY 1 agtgttgccactctgtgg 19  
||||||| |  
DB 2353 AGTGTGGCTCCTCTGG 2335

RESULT 11  
AAS01375/C  
ID AAS01375 standard; cDNA; 3114 BP.  
XX  
XX AAS01375;  
XX  
XX DT 04-JUL-2001 (first entry)  
XX  
XX Human TANGO 405 cDNA sequence.  
XX  
XX Human; TANGO 210; clone jthLal52h06; TANGO 364; TANGO 366; dectin-2;  
KW INTERCEPT 394; INTERCEPT 400; TANGO 405; cellular process regulator;  
KW gene therapy; growth modulator; proliferation; cell differentiation;  
KW lymphocyte; bone marrow cell migration; leukaemia; lymphoma;  
KW autoimmune disorder; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FH 154..783  
FT CDS  
FT /tag= a  
FT /product= "TANGO 405 protein"  
FT /note= "The ORF is specifically claimed"  
FT sig\_peptide 154..297  
FT mat\_peptide 298..780  
FT /tag= b  
FT /tag= c  
XX  
XX W0200118016-A1.  
PN  
XX  
XX 15-MAR-2001.  
XX  
XX 30-JUN-2000; 2000WO-US18174.  
PF

PA (MURA/) MURAMATSU T.  
 XX  
 DR WPI: 1996-406013/41.  
 DR P-PSDB: AAW06491.  
 XX  
 PT DNA sequence encoding beta-1,4-galactosyl:transferase-related  
 PT protein - useful for sterility diagnosis, and for assisting or  
 PT inhibiting fertilisation  
 XX  
 PS Claim 4; Page 7-9; 11pp; Japanese.  
 XX  
 CC The sequences given in AAT45081-82 represent two clones which encode  
 CC murine beta-1,4-galactosyltransferase-related proteins. The protein  
 CC can be used as a diagnostic agent for various diseases. It is esp.  
 CC useful in the diagnosis of sterility and in the aiding and  
 CC inhibiting of fertilisation. The cDNA's encoding the two beta-1,4-  
 CC galactosyltransferase-related proteins were isolated from F9 cancer  
 CC cells according to Huynh's method.  
 XX  
 SQ Sequence 3298 BP; 1007 A; 677 C; 758 G; 856 T; 0 other;  
 0;  
 Query Match 83.2%; Score 15.8; DB 17; Length 3298;  
 Best Local Similarity 89.5%; Pred. No. 1.7e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 agtgtggcactctctgtgg 19  
 |||||  
 Db 2909 agtgtggcactctctgtgg 2927  
 |||||  
 RESULT 13  
 AAQ13317/c  
 ID AAQ13317 standard; DNA; 4084 BP.  
 XX  
 AC AAQ13317;  
 XX  
 DT 17-DEC-2001 (updated)  
 DT 07-NOV-1991 (first entry)  
 XX  
 DE Duffy receptor gene.  
 KW Malaria; parasite; vaccine; ss.  
 XX  
 OS Plasmodium vivax Salvador I strain.  
 FH Key Location/Qualifiers  
 FT exon 1..3377  
 FT /\*tag= a  
 FT /number= 1  
 FT exon 3482..3561  
 FT /\*tag= b  
 FT /number= 2  
 FT exon 3781..3854  
 FT /\*tag= c  
 FT /number= 3  
 FT exon 3946..4084  
 FT /\*tag= d  
 FT /number= 4  
 FT CDS 227..3989  
 FT /\*tag= e  
 FT sig\_peptide 227..292  
 FT /\*tag= f  
 FT mat\_peptide 293..3986  
 FT /\*tag= g  
 XX  
 PN USN7554837-N.  
 XX  
 PD 23-JUL-1991.  
 XX  
 PF 20-JUL-1990; 90US-0554567.  
 XX  
 XX 20-JUL-1990; 90US-0554837.

XX (USSH ) NAT INST OF HEALTH.  
 PA Miller L, Adams J, Kaslow D, Xiangdong F;  
 PI  
 XX WPI: 1991-260184/35.  
 DR P-PSDB: AAR13457.  
 DR  
 XX Plasmodium Duffy receptor proteins - used in vaccines against  
 PT malaria and for producing antibodies which prevent malaria  
 PT infection.  
 XX  
 PS Disclosure; Fig 12; 63pp; English.  
 XX  
 CC The sequence was obtained from clones isolated from a P. vivax  
 CC genomic library using clone pCl1 encoding P. knowlesi Duffy  
 CC receptor gene as a probe. The existence of the 3' end introns was  
 CC determined by comparison with the corresponding gene from P.  
 CC knowlesi (AAQ13316). Unlike the P. knowlesi gene which was found to  
 CC have copies on 3 chromosomes, the P. vivax genome showed only one  
 CC copy of the gene. The gene can be used to express recombinant  
 CC Duffy receptor proteins for use in vaccines for malaria.  
 CC See also AAQ13316.  
 CC (Note: Revised entry submitted to correct the patent number format of  
 CC US Government-owned NTIS applications to prevent clashes with ongoing US  
 CC granted patent numbers. For further information please visit the Derwent  
 CC web site at www.derwent.com/dwpl/updates/ntis\_us.html.)  
 XX  
 SQ Sequence 4084 BP; 1562 A; 546 C; 851 G; 1125 T; 0 other;  
 0;  
 Query Match 83.2%; Score 15.8; DB 12; Length 4084;  
 Best Local Similarity 89.5%; Pred. No. 1.8e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 agtgtggcactctctgtgg 19  
 |||||  
 Db 1593 AGTTTTCACCTTCTGTGG 1575  
 |||||  
 RESULT 14  
 AAQ83524/c  
 ID AAQ83524 standard; DNA; 4084 BP.  
 XX  
 AC AAQ83524;  
 XX  
 DT 22-SEP-1995 (first entry)  
 XX  
 DE DABP gene.  
 XX  
 KW DABP; duffy antigen binding protein; binding domain; merozoite;  
 KW malaria; therapy; vaccine; ss.  
 XX  
 OS Plasmodium vivax.  
 XX  
 PN WO9507353-A.  
 XX  
 PD 16-MAR-1995.  
 XX  
 PF 07-SEP-1994; 94WO-US10230.  
 XX  
 PR 10-SEP-1993; 93US-0119677.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;  
 PI Wellens TE;  
 XX  
 DR WPI: 1995-123427/16.  
 DR P-PSDB: AAR70231.  
 XX  
 PT New erythrocyte binding domain polypeptide(s) - isolated from  
 PT Plasmodium binding proteins, used in diagnosis, treatment and



Mon May 20 10:14:45 2002

```
PT prevention of malaria
XX
PS
XX
XX Disclosure; Page 33-35; 8lpp; English.
XX
CC Sequences from the DABP gene (given in AAQ83524) were PCR amplified,
CC expressed on the surface of COS cells and tested for erythrocyte
CC binding to identify the binding domain polypeptide. A pref. DABP
CC binding domain comprises residues 1 to about 325 of the DABP protein
CC (AAAT70231). Recombinant binding domain was expressed in E. coli,
CC yeast, mammalian, insect, and in vaccinia virus and adenovirus-
CC infected cells. It provides protection against P. vivax.
XX
XX Sequence 4084 BP; 1562 A; 551 C; 846 G; 1125 T; 0 other;
XX
XX
XX Query Match 83.2%; Score 15.8; DB 16; Length 4084;
XX Best Local Similarity 89.5%; Pred. No. 1.8e+02;
XX Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 agtgttggaactctctgtgg 19
XX ||| || |||||
XX Db 1593 AGTTTGGCATTCTGTGG 1575
XX
XX RESULT 15
XX AAAT72889/c
XX ID AAT72889 standard; DNA; 4084 BP.
XX AC
XX AC AAT72889;
XX
XX DT 12-SEP-1997 (first entry)
XX
XX DE Duffy antigen binding protein coding sequence.
XX
XX DBL gene family: SAMP; stalic acid binding protein; vaccine; therapy;
XX Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
XX DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
XX Plasmodium; ss.
XX
XX OS Plasmodium vivax.
XX
XX FH Location/Qualifiers
XX FT key 227..3377
XX FT exon /*tag= a
XX FT /*number= 1
XX FT intron 3378..3482
XX FT /*tag= b
XX FT /*number= 1
XX FT exon 3483..3561
XX FT /*tag= c
XX FT /*number= 2
XX FT misc_feature 3521..3532
XX FT /*tag= d
XX FT /*note= "encode Leu-Leu-Leu"
XX FT intron 3562..3780
XX FT /*tag= e
XX FT /*number= 2
XX FT exon 3781..3854
XX FT /*tag= f
XX FT /*number= 3
XX FT intron 3855..3945
XX FT /*tag= g
XX FT /*number= 3
XX FT exon 3946..3989
XX FT /*tag= h
XX FT /*number= 4
XX
XX WO9640766-A2.
XX
XX PD 19-DEC-1996.
XX
XX PF 07-JUN-1996; 96WO-US09508.
XX
```

```
PR 07-JUN-1995; 95US-0487826.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
XX Wellem's TE;
XX
XX WPI; 1997-052231/05.
XX P-PSDB; AAM22478.
XX
XX New malaria vaccines - contains cysteine-rich DBL family protein
XX binding domains homologous domains of the Duffy and stalic acid
XX binding proteins
XX
XX Example 1; Page 31-33; 96pp; English.
XX
XX This sequence represents the full length coding sequence for the Duffy
XX antigen binding protein (DABP). DABP and the stalic acid binding protein
XX (SAMP) are soluble proteins that appear in the culture supernatant after
XX infected erythrocytes release merozoites. DABP and SAMP mediate the
XX binding of merozoites and schizonts to the erythrocyte surface. These
XX proteins are necessary for erythrocyte invasion by the parasite. This
XX sequence can be used in the compositions of the invention. The
XX compositions are for the treatment and prevention of malaria, and
XX comprise either a nucleotide sequence or encoded polypeptide of the
XX var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
XX genes having homology with conserved regions of DABP and SAMP. The
XX compositions are used for the treatment and prevention of malaria. They
XX are also used in the preparation of vaccines for inducing a protective
XX immune response in a mammal to Plasmodium merozoites (especially
XX Plasmodium falciparum or Plasmodium vivax).
XX
XX Sequence 4084 BP; 1562 A; 548 C; 849 G; 1125 T; 0 other;
XX
XX
XX Query Match 83.2%; Score 15.8; DB 18; Length 4084;
XX Best Local Similarity 89.5%; Pred. No. 1.8e+02;
XX Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 agtgttggaactctctgtgg 19
XX ||| || |||||
XX Db 1593 AGTTTGGCATTCTGTGG 1575
XX
XX Search completed: May 18, 2002, 17:46:00
XX Job time: 13749 sec
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:22:22 ; Search time 8624.33 seconds  
(without alignments)  
32.865 Million cell updates/sec

Title: US-09-623-329-25

Perfect score: 21

Sequence: 1 agcatggaggatgttgcaac 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST.\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_esti:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query %	Length	DB ID	Description
C 1	18.4	87.6	197	9	AW248227
C 2	18.4	87.6	363	10	BF047361
C 3	18.4	87.6	437	10	BE409003
C 4	18.4	87.6	440	10	BF923997
C 5	18.4	87.6	448	10	BF508210
C 6	18.4	87.6	507	10	BF340209
C 7	18.4	87.6	519	9	AW403936
C 8	18.4	87.6	536	10	BF125279
C 9	18.4	87.6	567	10	BF912509
C 10	18.4	87.6	588	10	BF751569
C 11	18.4	87.6	601	10	BE263559
C 12	18.4	87.6	620	10	BE276218
C 13	18.4	87.6	693	10	BF746885
C 14	18.4	87.6	696	10	BF026380
C 15	18.4	87.6	700	10	BE408733
C 16	18.4	87.6	715	10	BF474891
C 17	18.4	87.6	722	10	BE410079

C 18	18.4	87.6	723	10	BE409264
C 19	18.4	87.6	723	10	BF125476
C 20	18.4	87.6	768	10	BE264798
C 21	18.4	87.6	772	10	BE821144
C 22	18.4	87.6	822	10	BF423176
C 23	18.4	87.6	856	10	BE369126
C 24	18.4	87.6	873	10	BF437283
C 25	18.4	87.6	888	10	BE275187
C 26	18.4	87.6	916	10	BF973779
C 27	18.4	87.6	921	10	BF333629
C 28	18.4	87.6	924	10	BF394597
C 29	18.4	87.6	936	10	BF025671
C 30	18.4	87.6	974	10	BF476619
C 31	18.4	87.6	975	10	BE730159
C 32	18.4	87.6	1072	10	BE729323
C 33	17.8	84.8	520	10	BI060555
C 34	17.8	84.8	543	9	AV718920
C 35	17.8	84.8	579	12	AQ837888
C 36	17.8	84.8	588	9	AV719823
C 37	17.8	84.8	592	9	AV720612
C 38	17.8	84.8	595	9	AV720395
C 39	17.8	84.8	597	9	AV719439
C 40	17.8	84.8	597	9	AV719640
C 41	17.8	84.8	661	12	A2119639
C 42	17.8	84.8	664	12	A248566
C 43	17.8	84.8	989	12	CNS015X1
C 44	17.4	82.9	252	9	AA797617
C 45	17.4	82.9	293	12	BH122282

## ALIGNMENTS

RESULT 1

AW248227/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AW248227 197 bp mRNA linear EST 07-JAN-2000  
2819680.5prime NIH\_MGC\_7 Homo sapiens cDNA clone IMAGE:2819680 5',  
mRNA sequence.  
AW248227  
EST.  
GI:6591220  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 197)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Other-ESTs: 2819680.3prime  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mail.nih.gov  
Tissue procurement: DCTD/OTTP cDNA Library Preparation: Ling  
Hong/Kubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E.  
Consortium (LNL) DNA Sequencing by: Berkeley MGC sequencing  
project Clone distribution: MGC clone distribution information can  
be found through the I.M.A.G.E. Consortium/LNL at:  
www.bio.lnl.gov/brp/image/image.html Base Calling / Quality  
Scores: PHRED from University of Washington Genome Center. Vector  
Trimming: cross\_match from University of Washington Genome Center  
PHRAP suite. Poly-T identification: patmatch.pl from Berkeley  
Drosophila Genome Project. University of Washington Genome Center:  
http://www.genome.washington.edu  
Plate: L1CM2 row: D column: 17  
High quality sequence stop: 181.

FEATURES

source

Location/Qualifiers  
1..197  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="IMAGE:2819680"  
/clone\_lib="NIH\_MGC\_7"  
/tissue\_type="small cell carcinoma"  
/cell\_line="MGC3"

```

/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      40 a  54 c  68 g  35 t
ORIGIN

Query Match      87.6%; Score 18.4; DB 9; Length 197;
Best Local Similarity 95.0%; Pred. No. 4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2  gcatggagagatgttggcagc 21
      |||||||
Db  107 GCCTGGGAGATGTGGCAGC 88

RESULT  2
LOCUS   BF047361/c
DEFINITION  Bf047361 363 bp mRNA linear EST 11-OCT-2000
ACCESSION  IMAGE:3381697 5', mRNA sequence.
VERSION    BF047361
KEYWORDS   EST.
SOURCE     African clawed frog.
ORGANISM   Xenopus laevis
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
           Xenopodinae; Xenopus.
REFERENCE  1 (bases 1 to 363)
AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
           Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
           ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
           Waterston,R. and Wilson,R.
JOURNAL   WashU Xenopus EST project, 1999
COMMENT   Unpublished (1999)
           Contact: Sandy Clifton, Ph.D.
           WashU Xenopus EST project, 1999
           Washington University School of Medicine
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: est@watson.wustl.edu
           Library constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B.
           Gurdon (Wellcome/CRC Institute). DNA Sequencing by: Washington
           University Genome Sequencing Center
           Clone distribution: Xenopus clones from this library are available
           through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
           Seq primer: -40RP from Gibco.
           High quality sequence stop: 353.
           Location/Qualifiers
             1..363
               /organism="Xenopus laevis"
               /db_xref="taxon:8355"
               /clone="IMAGE:3381697"
               /clone_lib="Wellcome CRC psk animal cap"
               /tissue_type="embryo, animal cap"
               /lab_host="DH10B (phage-resistant)"
               /note="Vector: pBluescript SK-; Site_1: NotI; Site_2:
               EcoRI; cDNAs were oligo-dT primed and directionally
               cloned. Staging according to Nieuwkoop and Faber. Library
               was constructed by N. Garrett, P. LeMaire, A.M. Zorn, and
               J.B. Gurdon (Wellcome/CRC Institute). Note: This is a
               Xenopus Gene Collection (XGC) library."
BASE COUNT      86 a  96 c  114 g  67 t
ORIGIN

/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      40 a  54 c  68 g  35 t
ORIGIN

Query Match      87.6%; Score 18.4; DB 9; Length 197;
Best Local Similarity 95.0%; Pred. No. 4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2  gcatggagagatgttggcagc 21
      |||||||
Db  107 GCCTGGGAGATGTGGCAGC 88

RESULT  2
LOCUS   BF047361/c
DEFINITION  Bf047361 363 bp mRNA linear EST 11-OCT-2000
ACCESSION  IMAGE:3381697 5', mRNA sequence.
VERSION    BF047361
KEYWORDS   EST.
SOURCE     African clawed frog.
ORGANISM   Xenopus laevis
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
           Xenopodinae; Xenopus.
REFERENCE  1 (bases 1 to 363)
AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
           Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
           ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
           Waterston,R. and Wilson,R.
JOURNAL   WashU Xenopus EST project, 1999
COMMENT   Unpublished (1999)
           Contact: Sandy Clifton, Ph.D.
           WashU Xenopus EST project, 1999
           Washington University School of Medicine
           4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
           Tel: 314 286 1800
           Fax: 314 286 1810
           Email: est@watson.wustl.edu
           Library constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B.
           Gurdon (Wellcome/CRC Institute). DNA Sequencing by: Washington
           University Genome Sequencing Center
           Clone distribution: Xenopus clones from this library are available
           through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
           Seq primer: -40RP from Gibco.
           High quality sequence stop: 353.
           Location/Qualifiers
             1..363
               /organism="Xenopus laevis"
               /db_xref="taxon:8355"
               /clone="IMAGE:3381697"
               /clone_lib="Wellcome CRC psk animal cap"
               /tissue_type="embryo, animal cap"
               /lab_host="DH10B (phage-resistant)"
               /note="Vector: pBluescript SK-; Site_1: NotI; Site_2:
               EcoRI; cDNAs were oligo-dT primed and directionally
               cloned. Staging according to Nieuwkoop and Faber. Library
               was constructed by N. Garrett, P. LeMaire, A.M. Zorn, and
               J.B. Gurdon (Wellcome/CRC Institute). Note: This is a
               Xenopus Gene Collection (XGC) library."
BASE COUNT      86 a  96 c  114 g  67 t
ORIGIN

/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      40 a  54 c  68 g  35 t
ORIGIN

Query Match      87.6%; Score 18.4; DB 10; Length 363;
Best Local Similarity 95.0%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2  gcatggagagatgttggcagc 21
      |||||||
Db  74 GCATGGGAAATGTGGCAGC 55

RESULT  3
LOCUS   BE409003/c
DEFINITION  Bf409003 437 bp mRNA linear EST 21-JUL-2000
ACCESSION  BE409003
VERSION    BE409003.1 GI:9345453
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 437)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-remail.nih.gov
           Tissue procurement: ATCC
           cDNA Library Preparation: Ling Hong/Rubin Laboratory
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
           Plate: LLCM337 row: h column: 08
           High quality sequence stop: 437.
           Location/Qualifiers
             1..437
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:3637687"
               /clone_lib="NIH_MGC_21"
               /tissue_type="Choriocarcinoma"
               /lab_host="DH10B (phage-resistant)"
               /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
               Site_2: EcoRI; cDNA made by oligo-dT priming.
               Directionally cloned into EcoRI/XhoI sites using the
               following 5' adaptor: GGCACGAG(G). Size-selected >500bp
               for average insert size 1.8kb. Library constructed by
               Ling Hong in the laboratory of Gerald M. Rubin (University
               of California, Berkeley) using ZAP-cDNA synthesis kit
               (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      94 a  144 c  141 g  58 t
ORIGIN

Query Match      87.6%; Score 18.4; DB 10; Length 437;
Best Local Similarity 95.0%; Pred. No. 5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2  gcatggagagatgttggcagc 21
      |||||||
Db  351 GCCTGGGAGATGTGGCAGC 332

RESULT  4
LOCUS   BF923997/c
DEFINITION  Bf923997 440 bp mRNA linear EST 19-JAN-2001
ACCESSION  BF923997
VERSION    BF923997.1 GI:12319885
KEYWORDS   EST.
SOURCE     human.

```

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS 1 (bases 1 to 440)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baid,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&t2=CM2-NT0170-  
251100-330-e05&t3=2000-11-25&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 6  
High quality sequence stop: 440.

FEATURES  
source  
1..440  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NT0170"  
/dev\_stage="Adult"  
/note="Organ: nervous\_tumor; Vector: puc18; Site\_1: Sma1;  
Site\_2: Sma1; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 93 a 140 c 130 g 77 t  
ORIGIN  
Query Match 87.6%; Score 18.4; DB 10; Length 440;  
Best Local Similarity 95.0%; Pred. No. 5e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcatgggagatgttgcgcg 21  
||||| ||||||| |||||||  
Db 217 CCCTGGGAGATGTTGGCAGC 198

RESULT 5  
BF508210/c  
LOCUS UI-H-B14-apy-e-11-0-UI.sl NCI\_CGAP\_Sub8 Homo sapiens cDNA clone  
DEFINITION IMAGE:3089012 3', mRNA sequence.  
ACCESSION BF508210 448 bp mRNA linear EST 06-DEC-2000  
VERSION BF508210.1 GI:11591508  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 448)  
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:  
NCI-CCAP clone distribution information can be found through the  
I.M.A.G.E. Consortium/LLNL at:  
www.bio.lnl.gov/bbrp/image/image.html The following repetitive  
elements were found in this cDNA sequence: 285-389,  
>RMR20#Unknown/LTR?

Seq primer: M13 Forward

POLYA=Yes.

FEATURES  
source

Location/Qualifiers  
1..448  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3089012"  
/clone\_lib="NCI\_CGAP\_Sub8"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not 1; Site\_2: Eco RI; NCI\_CGAP\_Sub8  
is a subtracted library derived from NCI\_CGAP\_Sub5. The  
NCI\_CGAP\_Sub8 library had 2.5 million recombinants. A  
single-stranded DNA preparation of NCI\_CGAP\_Sub5 was used  
as a tracer in a subtractive hybridization with a driver  
comprising: a pool of clones from NCI\_CGAP\_Sub5 (IMAGE  
clone ids 2732833-2737415, 3068040-3069191); 25% of the  
driver population), a pool of clones from NCI\_CGAP\_Sub4  
(IMAGE clone ids 2723592-2729326; 25% of the driver  
population), NCI\_CGAP\_Sub6 (pool AIF-AJU, IMAGE ids  
2728969-2733190; 25% of the driver population), and  
NCI\_CGAP\_Sub7 (IMAGE ids 3069192-3072238, 3081864-3084550  
); 25% of the driver population). Subtraction was  
performed as previously described (Bonaldo, Lennon &  
Soares (1996): Normalization and Subtraction: Genome Research  
Approaches To Facilitate Gene Discovery. 6, 791-806.  
TAG\_LIB=NCI\_CGAP\_Br2  
TAG\_TISSUE=breast  
TAG\_SEQ=AAACC"

BASE COUNT 139 a 102 c 108 g 99 t  
ORIGIN

Query Match 87.6%; Score 18.4; DB 10; Length 448;  
Best Local Similarity 95.0%; Pred. No. 5e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcatgggagatgttgcgcg 21  
||||| ||||||| |||||||  
Db 385 GCATGGTATGATGTTGGCAGC 366

RESULT 6  
BG340209/c  
LOCUS BG340209 507 bp mRNA linear EST 27-FEB-2001  
DEFINITION 602437792F1 NIH\_MGC\_46 Homo sapiens cDNA clone IMAGE:455462 5',  
mRNA sequence.

ACCESSION BG340209  
VERSION BG340209.1 GI:13146636  
KEYWORDS EST.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 507)  
AUTHORS NIH-MGC http://mgs.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory



cdNA Library Preparation: Ling Hong/Rubin Laboratory  
cdNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov h column: 20  
Plate: LLCMI753 row: a column: 19  
High quality sequence stop: 561.  
Location/Qualifiers  
1. 567  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4938528"  
/clone\_lib="NCI\_CGAP\_Brn67"  
/tissue\_type="anaplastic oligodendroglioma with lp/19q  
loss"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: brain; Vector: pCMV-SPORT6; Site:1: NotI;  
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2.3 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."  
Note: this is a NCI\_CGAP Library. 1\*  
BASE COUNT 99 a 188 c 169 g 111 t  
ORIGIN  
Query Match 87.6%; Score 18.4; DB 10; Length 567;  
Best Local Similarity 95.0%; Pred. No. 5.4e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 gcatgggagatgttgcagc 21  
||| ||||| ||||| |||||  
Db 60 GCCTGGGAGATGTTGCAGC 41  
RESULT 10  
BG751569/c  
LOCUS BG751569.1 GI:14062222  
DEFINITION Homo sapiens  
mRNA sequence.  
EST.  
ACCESSION BG751569  
VERSION BG751569.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 588)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: David N. Louis, M.D.  
cdNA Library Preparation: Life Technologies, Inc.  
cdNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov h column: 01  
Plate: LLAM10874 row: b column: 01  
High quality sequence stop: 561.  
Location/Qualifiers  
1. 567  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4938528"  
/clone\_lib="NCI\_CGAP\_Brn67"  
/tissue\_type="anaplastic oligodendroglioma with lp/19q  
loss"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: brain; Vector: pCMV-SPORT6; Site:1: NotI;  
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2.3 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."  
Note: this is a NCI\_CGAP Library. 1\*  
BASE COUNT 99 a 188 c 169 g 111 t  
ORIGIN  
Query Match 87.6%; Score 18.4; DB 10; Length 567;  
Best Local Similarity 95.0%; Pred. No. 5.4e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 gcatgggagatgttgcagc 21  
||| ||||| ||||| |||||  
Db 60 GCCTGGGAGATGTTGCAGC 41  
RESULT 10  
BG751569/c  
LOCUS BG751569.1 GI:14062222  
DEFINITION Homo sapiens  
mRNA sequence.  
EST.  
ACCESSION BG751569  
VERSION BG751569.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 588)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC

cdNA Library Preparation: Ling Hong/Rubin Laboratory  
cdNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov h column: 20  
Plate: LLCMI753 row: a column: 19  
High quality sequence stop: 586.  
Location/Qualifiers  
1. 588  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4874179"  
/clone\_lib="NIH\_MGC\_43"  
/tissue\_type="normal pigmented retinal epithelium"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pORF7; Site:1: XhoI; Site:2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GCACGAG(G). Library constructed by Ling Hong  
in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library. 1\*  
BASE COUNT 128 a 187 c 196 g 77 t  
ORIGIN  
Query Match 87.6%; Score 18.4; DB 10; Length 588;  
Best Local Similarity 95.0%; Pred. No. 5.5e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 gcatgggagatgttgcagc 21  
||| ||||| ||||| |||||  
Db 568 GCCTGGGAGATGTTGCAGC 549  
RESULT 11  
BE263559/c  
LOCUS BE263559.1 GI:9137102  
DEFINITION Homo sapiens  
mRNA sequence.  
EST.  
ACCESSION BE263559  
VERSION BE263559.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 601)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Plate: LLCM212 row: a column: 19  
High quality sequence stop: 601.  
Location/Qualifiers  
1. 601  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3533850"  
/clone\_lib="NIH\_MGC\_7"  
/tissue\_type="small cell carcinoma"  
/cell\_line="MGC3"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: pORF7; Site:1: XhoI; Site:2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit



DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LLCM827 row: o column: 08  
 High quality sequence stop: 692.  
 Location/Qualifiers

#### FEATURES

source  
 1..696  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3954655"  
 /clone\_lib="NIH\_MGC\_20"  
 /tissue\_type="melanotic melanoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: skin; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 135 a 218 c 210 g 133 t  
 ORIGIN

Query Match 87.6%; Score 18.4; DB 10; Length 696;  
 Best Local Similarity 95.0%; Pred. No. 5.7e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcatgggagatgttggcagc 21  
 |||||  
 Db 206 GCCTGGGAGATGTTGGCAGC 187

RESULT 15  
 BE408733/c  
 LOCUS  
 DEFINITION BE408733 700 bp mRNA linear EST 21-JUL-2000  
 601303415F1 NIH\_MGC\_21 Homo sapiens cDNA clone IMAGE:3638186 5',  
 mRNA sequence.  
 ACCESSION BE408733  
 VERSION BE408733.1 GI:9345183  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (bases 1 to 700)  
 NIH-MGC http://mgc.nci.nih.gov/.  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs@mail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LLCM338 row: m column: 03  
 High quality sequence stop: 652.  
 Location/Qualifiers  
 1..700  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3638186"  
 /clone\_lib="NIH\_MGC\_21"  
 /tissue\_type="Choriocarcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: placenta; Vector: pOTB7; Site\_1: XhoI;  
 Site\_2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by

#### FEATURES

source  
 1..700  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3638186"  
 /clone\_lib="NIH\_MGC\_21"  
 /tissue\_type="Choriocarcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: placenta; Vector: pOTB7; Site\_1: XhoI;  
 Site\_2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by

Ling Hong in the laboratory of Gerald M. Rubin (University  
 of California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 138 a 211 c 213 g 138 t  
 ORIGIN

Query Match 87.6%; Score 18.4; DB 10; Length 700;  
 Best Local Similarity 95.0%; Pred. No. 5.7e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 gcatgggagatgttggcagc 21  
 |||||  
 Db 207 GCCTGGGAGATGTTGGCAGC 188

Search completed: May 18, 2002, 17:22:25  
 Job time: 14696 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:26:41 ; Search time 250.21 Seconds  
(without alignments)  
20.616 Million cell updates/sec

Title: US-09-623-329-25

Perfect score: 21

Sequence: 1 agcatgggagatgttggcgc 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, NA.\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*

4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*

5: /cgn2\_6/ptodata/2/ina/PTCTUS\_COMB.seq.\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	17.8	84.8	7011	4	US-09-268-163-9
2	16.8	80.0	2840	5	PCT-US93-06251-64
3	16.8	80.0	2875	5	PCT-US93-06251-63
C 4	15.8	75.2	6294	4	US-08-976-259-5
5	15.2	72.4	2365	4	US-09-183-706-42
6	15.2	72.4	2365	4	US-09-567-995-42
7	15.2	72.4	10348	2	US-08-457-273B-41
8	15.2	72.4	10348	3	US-08-556-419-13
9	15.2	72.4	10348	4	US-09-041-886-14
10	15.2	72.4	10366	1	US-08-246-982A-5
11	15.2	72.4	10366	1	US-08-453-265-5
C 12	14.8	70.5	74	4	US-08-194-560-5
13	14.8	70.5	1245	3	US-09-242-050-1
14	14.8	70.5	1417	1	US-08-713-828-2
15	14.8	70.5	1417	2	US-08-919-627-2
16	14.8	70.5	1417	2	US-09-096-245-2
C 17	14.8	70.5	3804	2	US-08-483-488-5
C 18	14.6	69.5	272	4	US-09-318-978-16
C 19	14.6	69.5	586	4	US-09-001-141-3
20	14.6	69.5	1392	4	US-09-163-444-1
21	14.6	69.5	1496	1	US-08-601-435-1
22	14.6	69.5	1496	2	US-08-931-047-1
23	14.6	69.5	1496	2	US-08-783-202-1
C 24	14.6	69.5	1662	1	US-08-336-408B-5
C 25	14.6	69.5	1662	5	PCT-US91-00399-5
26	14.6	69.5	1735	4	US-09-163-444-3
27	14.6	69.5	4054	1	US-07-618-946B-2

C 28	14.6	69.5	4315	3	US-08-882-046-3	Sequence 3, Appli
C 29	14.6	69.5	4847	3	US-09-045-632-47	Sequence 47, Appli
C 30	14.6	69.5	4847	3	US-09-045-632-48	Sequence 48, Appli
C 31	14.6	69.5	6830	2	US-08-822-445-1	Sequence 1, Appli
C 32	14.6	69.5	6830	4	US-09-396-540-1	Sequence 1, Appli
C 33	14.6	69.5	29604	3	US-08-781-891-207	Sequence 207, App
C 34	14.6	69.5	152331	3	US-09-128-155-16	Sequence 16, Appli
C 35	14.4	68.6	72	3	US-09-046-247-130	Sequence 130, App
C 36	14.4	68.6	1400	3	US-08-695-191-14	Sequence 14, Appli
C 37	14.4	68.6	1400	3	US-08-682-080-14	Sequence 13, Appli
C 38	14.4	68.6	1434	3	US-08-695-191-13	Sequence 13, Appli
C 39	14.4	68.6	1434	3	US-08-682-080-13	Sequence 11, Appli
C 40	14.4	68.6	1495	4	US-09-364-230-11	Sequence 34, Appli
C 41	14.4	68.6	1524	1	US-08-197-792-34	Sequence 34, Appli
C 42	14.4	68.6	1524	1	US-08-459-850-34	Sequence 34, Appli
C 43	14.4	68.6	1524	1	US-08-459-214-34	Sequence 1, Appli
C 44	14.4	68.6	5238	3	US-09-080-855-1	Sequence 5, Appli
C 45	14.2	67.6	308	1	US-08-627-706-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1

US-09-268-163-9/C

; Sequence 9, Application US/09268163B

; Patent No. 6353091

; GENERAL INFORMATION:

; APPLICANT: Lipscombe, Diane

; TITLE OF INVENTION: HUMAN N-TYPE CALCIUM CHANNEL ISOFORM AND USES THEREOF

; FILE REFERENCE: B1055/7000

; CURRENT APPLICATION NUMBER: US/09/268.163B

; CURRENT FILING DATE: 1999-03-12

; EARLIER APPLICATION NUMBER: US 60/077.901

; EARLIER FILING DATE: 1998-03-13

; NUMBER OF SEQ ID NOS: 28

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 9

; LENGTH: 7011

; TYPE: DNA

; ORGANISM: Rattus norvegicus

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..7008

US-09-268-163-9

Query Match 84.8%; Score 17.8; DB 4; Length 7011;

Best Local Similarity 90.5%; Pred. No. 5.9;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 agcatgggagatgttggcgc 21

||||| ||||||| |||||||

Db 2268 AGCAATGGAGATGTGGCAGC 2248

RESULT 2

PCT-US93-06251-64

; Sequence 64, Application PC/TUS9306251

; GENERAL INFORMATION:

; APPLICANT: Wickstrom, Eric and Rife, Jason P.

; TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing

; NUMBER OF SEQUENCES: 93

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER

; STREET: 400 Garden City Plaza

; CITY: Garden City

; STATE: NY

; COUNTRY: USA

; ZIP: 11530

; COMPUTER READABLE FORM:



```

RESULT 7
US-08-457-273B-41
: Sequence 41, Application US/08457273B
: Patent No. 5849995
: GENERAL INFORMATION:
: APPLICANT: Hayden, Michael
: APPLICANT: Lin, Biaoyang
: APPLICANT: Nasir, Jamal
: TITLE OF INVENTION: Mouse Model for Huntington's Disease and
: TITLE OF INVENTION: Related DNA Sequences
: NUMBER OF SEQUENCES: 42

```

```

Query Match          72.4%;   Score 15.2;   DB 3;   Length 10348;
Best Local Similarity 85.0%;   Pred. No. 1.2e+02;
Matches 17;   Conservative 0;   Mismatches 3;   Indels 0;   Gaps 0;

Qy 1 agcatggggagatgttggcag 20
    ||| ||| ||| ||| ||| ||| ||| |||

```

Db 7395 agcctgtgagatgttgcag 7414

## RESULT 9

US-09-041-886-14  
; Sequence 14, Application US/09041886  
; Patent No. 6235872  
; GENERAL INFORMATION:  
; APPLICANT: Bredesen, Dale E.  
; APPLICANT: Rabizadeh, Shiroz  
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence  
; NUMBER OF SEQUENCES: 72  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/041,886  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 2626  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10348 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 316..9748  
; US-09-041-886-14

Query Match 72.4%; Score 15.2; DB 4; Length 10348;  
Best Local Similarity 85.0%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 agcatggagatgttgcag 20  
||| || ||||| |||||  
Db 7395 AGCCTGTGAGATGTGTCAG 7414

## RESULT 10

US-08-246-982A-5  
; Sequence 5, Application US/08246982A  
; Patent No. 5686288  
; GENERAL INFORMATION:  
; APPLICANT: MacDonald, Marcy E.  
; APPLICANT: Ambrose, Christine M.  
; APPLICANT: Duyao, Mabel P.  
; APPLICANT: Gusella, James F.  
; TITLE OF INVENTION: Huntingtin DNA, Protein And Uses Thereof  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
; STREET: 1100 New York Avenue  
; CITY: Washington

; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/246,982A  
; FILING DATE: May 20, 1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldstein, Jorge, A.  
; REGISTRATION NUMBER: 29,021  
; REFERENCE/DOCKET NUMBER: 0609.3880002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10366 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 316..9748  
; US-08-246-982A-5

Query Match 72.4%; Score 15.2; DB 1; Length 10366;  
Best Local Similarity 85.0%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 agcatggagatgttgcag 20  
||| || ||||| |||||  
Db 7395 AGCCTGTGAGATGTGTCAG 7414

## RESULT 11

US-08-453-265-5  
; Sequence 5, Application US/08453265  
; Patent No. 5693757  
; GENERAL INFORMATION:  
; APPLICANT: MacDonald, Marcy E.  
; APPLICANT: Ambrose, Christine M.  
; APPLICANT: Duyao, Mabel P.  
; APPLICANT: Gusella, James F.  
; TITLE OF INVENTION: Huntingtin DNA, Protein And Uses Thereof  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
; STREET: 1100 New York Avenue  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,265  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, Steven R.  
; REGISTRATION NUMBER: 36,203  
; REFERENCE/DOCKET NUMBER: 0609.3880003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10366 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 316...9748  
US-08-453-265-5

Query Match 72.4%; Score 15.2; DB 1; Length 10366;  
Best Local Similarity 85.0%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 agcatggagatgttgacg 20  
||| || ||||| |||||  
Db 7395 AGCCTGTGAGATGTGGCAG 7414

RESULT 12  
US-08-194-560-5/C

Sequence 5, Application US/08194560  
Patent No. 6255062  
GENERAL INFORMATION:  
APPLICANT: Campbell, Judith L.  
APPLICANT: Budd, Martin E.  
TITLE OF INVENTION: B-Type DNA Polymerases  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Honbach, Test, Albritton & Herbert  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/194,560  
FILING DATE: 14-FEB-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Trecartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-59515/RFT/RMS  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 74 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-194-560-5  
Query Match 70.5%; Score 14.8; DB 4; Length 74;  
Best Local Similarity 88.9%; Pred. No. 91;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 catggagatgttgacg 20  
||||| ||||| |||||  
Db 25 CATGGAGATCTGTGCAG 8

RESULT 13  
US-09-242-050-1/c  
Sequence 1, Application US/09242050  
Patent No. 6136560  
GENERAL INFORMATION:  
APPLICANT: SmithKline Beecham Corporation  
APPLICANT: Black, Michael T.  
APPLICANT: Lawlor, Elizabeth J.  
TITLE OF INVENTION: No. 6136560el Compounds  
FILE REFERENCE: GM50022  
CURRENT APPLICATION NUMBER: US/09/242,050  
CURRENT FILING DATE: 1999-02-08  
EARLIER APPLICATION NUMBER: 60/041,130  
EARLIER FILING DATE: 1997-03-20  
EARLIER APPLICATION NUMBER: 60/041,131  
EARLIER FILING DATE: 1997-03-20  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 1245  
TYPE: DNA  
ORGANISM: Streptococcus pneumoniae  
US-09-242-050-1

Query Match 70.5%; Score 14.8; DB 3; Length 1245;  
Best Local Similarity 88.9%; Pred. No. 1.4e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 atgggagatgttgacg 21  
| ||||| ||||| ||  
Db 597 AAGCGAGATCTGGCATC 580

RESULT 14  
US-08-713-828-2  
Sequence 2, Application US/08713828  
Patent No. 5683910  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goll, Surya K.  
TITLE OF INVENTION: NOVEL HUMAN PHOSPHORYLASE  
TITLE OF INVENTION: KINASE GAMMA SUBUNIT  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/713,828  
FILING DATE: Filed Herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0068 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1417 base pairs

Mon May 20 10:14:50 2002

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE: Consensus
; US-08-713-828-2

```

```

Query Match 70.5%; Score 14.8; DB 1; Length 1417;
Best Local Similarity 88.9%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 4 atgggagatgttggcagc 21
| | | | | | | | | | |
Db 1282 AAGGGAGATGTTGGCACC 1299

```

```

RESULT 15
US-08-919-627-2
; Sequence 2, Application US/08919627
; Patent No. 5833981
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goll, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN PHOSPHORYLASE
; TITLE OF INVENTION: KINASE GAMMA SUBUNIT
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/919,627
; FILING DATE: August 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/713,828
; FILING DATE: September 13, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0068-1 DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1417 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE: Consensus
; US-08-919-627-2

```

```

Query Match 70.5%; Score 14.8; DB 2; Length 1417;
Best Local Similarity 88.9%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 4 atgggagatgttggcagc 21
| | | | | | | | | | |

```

Db 1282 AAGGGAGATGTTGGCACC 1299

Search completed: May 18, 2002, 17:26:43  
Job time: 12845 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:46:00 ; Search time 1024.22 Seconds  
(without alignments)  
35.203 Million cell updates/sec

Title: US-09-623-329-25

Perfect score: 21

Sequence: 1 agcatggagatgttgcagc 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

1: /SID55/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID55/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID55/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID55/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID55/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID55/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID55/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID55/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID55/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID55/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID55/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID55/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID55/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID55/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID55/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID55/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	21	AAZ11675	Oligo specific for
2	18.4	87.6	32167	20 ABA20621	Human nervous syst
3	17.4	82.9	552	22 ABA63472	Human foetal liver
4	17.4	82.9	552	22 ABA30670	Probe #9136 for ge
5	17.4	82.9	552	22 AAK12005	Human brain expres
6	17.4	82.9	552	22 AAK37711	Human bone marrow
7	17.4	82.9	552	22 AAI18467	Probe #8400 for ge
8	17.4	82.9	89047	22 AAF28547	Genomic fragment #
9	16.8	80.0	2839	11 AAQ03742	Human SKI related

10	16.8	80.0	2888	11 AAQ03743	Human SKI related
11	16.4	78.1	1026	23 AAS66199	DNA encoding novel
12	16.2	77.1	700	22 AAH92223	Human inflammatory
13	16.2	77.1	700	22 AAH92224	Human inflammatory
14	16.2	77.1	935	21 AAC46188	Arabidopsis thalia
15	16.2	77.1	938	21 AAC35293	Arabidopsis thalia
16	16.2	77.1	2637	22 AAH77490	Human zinc finger
17	16.2	77.1	8246	23 AAS59621	Propionibacterium
18	16.2	77.1	16738	22 AAK70864	Human immune/haema
19	16.2	77.1	25871	21 AAA09888	Human genomic OCTN
20	15.8	75.2	215	22 ABA88796	Escherichia coli p
21	15.8	75.2	931	22 AAS25834	Human cDNA encodin
22	15.8	75.2	1284	22 AAS26295	Human cDNA encodin
23	15.8	75.2	1877	21 AAZ94945	Human carboxydrate
24	15.8	75.2	1895	22 ABA83001	Human transcriptio
25	15.8	75.2	2090	21 AAC51803	Arabidopsis thalia
26	15.8	75.2	2327	22 AAI60660	Human polynucleoti
27	15.8	75.2	2334	22 AAI58874	Human polynucleoti
28	15.8	75.2	2400	20 AAV80666	Human DNAX toll-11
29	15.8	75.2	2400	20 AAD26283	Human DNAX toll-11
30	15.8	75.2	2500	22 AAH14072	Human cDNA sequenc
31	15.8	75.2	2883	23 ABL25994	Drosophila melanog
32	15.8	75.2	3159	23 ABL26017	Drosophila melanog
33	15.8	75.2	4611	23 ABL16081	Drosophila melanog
34	15.8	75.2	4865	20 AAV80675	Human DNAX toll-11
35	15.8	75.2	4865	20 AAD26292	Human DNAX toll-11
36	15.8	75.2	4865	24 AAD26306	Human DNAX toll-11
37	15.8	75.2	5378	22 AAS27810	DNA encoding novel
38	15.8	75.2	5409	23 ABL26016	Drosophila melanog
39	15.8	75.2	6294	19 AAV31191	E. coli J96 pathog
40	15.8	75.2	6306	22 ABA15649	Human nervous syst
41	15.8	75.2	6695	22 ABA15648	Human nervous syst
42	15.8	75.2	6697	22 ABA15650	Human nervous syst
43	15.8	75.2	6786	22 AAC84836	Partial nucleotide
44	15.8	75.2	10665	22 AAC84826	Partial nucleotide
45	15.8	75.2	11756	23 ABL16080	Drosophila melanog

## ALIGNMENTS

RESULT 1

AAZ11675

ID AAZ11675 standard; DNA; 21 BP.

XX AAZ11675;

AC AAZ11675;

XX 19-NOV-1999 (first entry)

DE Oligo specific for EBV BARRF-1 RNA.

XX Epstein Barr Virus; EBV infection; viral; gene transcription; EBEB-1;

XX Epstein Barr early RNA; Epstein Barr nuclear antigen 1; EBNA-1; LMP-1;

XX latent membrane protein; LMP-2; VIL10; BCRF-1; BARRF1; BDLF2; NASBA;

XX EBV-associated malignancy; primer; ss.

XX Synthetic.

OS Epstein-barr virus.

XX WO9945155-A2.

XX 10-SEP-1999.

PD 01-MAR-1999; 99WO-EP01392.

XX 04-MAR-1998; 98EP-0200655.

XX 14-DEC-1998; 98EP-0204231.

PA (ALKU ) AKZO NOBEL NV.

XX Vervoort MBHJ, Van Den Brule AJC, Middelorp JM;

XX WPI; 1999-551051/46.





PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 08-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI; 2001-541565/60.  
Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
useful for preventing, diagnosing and/or treating nervous system  
cancers and metastases -  
Disclosure; SEQ ID NO 12952; 1701pp + Sequence Listing; English.  
The invention relates to novel genes (AB11004-AB21534) and proteins  
(AB114678-AB18001) useful for preventing, treating or ameliorating  
medical conditions e.g. by protein or gene therapy. The genes are  
isolated from a range of human tissues disclosed in the specification.  
The nucleic acids, proteins, antibodies and (ant)agonists are useful  
in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
and ovarian cancer and other cancers of the adrenal gland, bone, bone  
marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
(b) immune disorders e.g. Addison's disease, allergies, autoimmune  
haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
colitis; (c) cardiovascular disorders such as myocardial ischaemias;

CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 32167 BP; 7272 A; 8300 C; 8385 G; 8210 T; 0 other;  
Query Match 87.6%; Score 18.4; DB 22; Length 32167;  
Best Local Similarity 95.0%; Pred. No. 32;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 gcatggagatgtggcagc 21  
|| |||||  
Db 29226 GCCTGGGAGATGTGGCAGC 29207  
RESULT 3  
ABA63472  
ID ABA63472 standard; DNA; 552 BP.  
XX  
AC ABA63472;  
XX  
DT 01-FEB-2002 (first entry)  
XX Human foetal liver single exon nucleic acid probe #11777.  
XX Human; foetal liver; gene expression; single exon nucleic acid probe: ss.  
XX Homo sapiens.  
XX W0200157277-A2.  
XX 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-0500669.  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
(MOLK-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel UK, Chen W, Rank DR;  
XX WPI; 2001-483447/52.  
XX Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human fetal liver -  
XX  
PS Claim 1; SEQ ID NO 11777; 639pp + sequence listing; English.  
XX  
CC The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC foetal liver. The present sequence is a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 552 BP; 163 A; 95 C; 201 G; 93 T; 0 other;  
Query Match 82.9%; Score 17.4; DB 22; Length 552;  
Best Local Similarity 94.7%; Pred. No. 60;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 agcatggagatgttgcca 19  
||||| |||||||||  
Db 453 agcatggagatgttgcca 471

## RESULT 4

ABA30670  
ID ABA30670 standard; DNA: 552 BP.

XX  
AC ABA30670;

XX  
DT 23-JAN-2002 (first entry)

XX  
DE Probe #9136 for gene expression analysis in human heart cell sample.

XX  
KW Human; gene expression; heart; microarray; vascular system; probe;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease; ss.

OS Homo sapiens.

XX  
XX WO200157274-A2.

PN  
XX 09-AUG-2001.

XX  
PD 30-JAN-2001; 2001WO-US00666.

XX  
PF 04-FEB-2000; 2000US-0180312.

XX  
PR 26-MAY-2000; 2000US-0207456.

XX  
PR 30-JUN-2000; 2000US-0608408.

XX  
PR 03-AUG-2000; 2000US-0632366.

XX  
PR 21-SEP-2000; 2000US-0234687.

XX  
PR 27-SEP-2000; 2000US-0236359.

XX  
PR 04-OCT-2000; 2000GB-0024263.

XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.

XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX  
DR WPI; 2001-488899/53.

XX  
PT Single exon nucleic acid probes for analyzing gene expression in human

XX  
PS hearts -

XX  
PS Claim 1; SEQ ID No 9136; 530pp; English.

XX  
CC The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart. The  
CC present sequence is one such probe. The probes may be used for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from the human heart via microarrays. By measuring gene expression, the  
CC probes are useful for predicting, diagnosing, grading, staging,  
CC monitoring and prognosing diseases of the human heart and vascular system  
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
CC congenital heart disease.

CC  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

XX  
SQ Sequence 552 BP; 163 A; 95 C; 201 G; 93 T; 0 other;

Query Match 82.9%; Score 17.4; DB 22; Length 552;

Best Local Similarity 94.7%; Pred. No. 60;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 agcatggagatgttgcca 19  
||||| |||||||||

Db 453 agcatggagatgttgcca 471

## RESULT 5

AAK12005

ID AAK12005 standard; DNA: 552 BP.

XX  
AC AAK12005;

XX  
DT 05-NOV-2001 (first entry)

XX  
DE Human brain expressed single exon probe SEQ ID NO: 11996.

XX  
KW Human; brain expressed exon; gene expression analysis; probe;

XX  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

XX  
KW epilepsy; cancer; ss.

XX  
OS Homo sapiens.

XX  
PN WO200157275-A2.

XX  
PD 09-AUG-2001.

XX  
PF 30-JAN-2001; 2001WO-US00667.

XX  
PR 04-FEB-2000; 2000US-0180312.

XX  
PR 26-MAY-2000; 2000US-0207456.

XX  
PR 30-JUN-2000; 2000US-0608408.

XX  
PR 03-AUG-2000; 2000US-0632366.

XX  
PR 21-SEP-2000; 2000US-0234687.

XX  
PR 27-SEP-2000; 2000US-0236359.

XX  
PR 04-OCT-2000; 2000GB-0024263.

XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.

XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX  
DR WPI; 2001-483446/52.

XX  
PT Single exon nucleic acid probes for analyzing gene expression in human

XX  
PT brains -

XX  
PS Example 4; SEQ ID NO: 11996; 650pp + Sequence Listing; English.

XX  
CC The present invention provides a number of single exon nucleic acid

XX  
CC probes which are derived from genomic sequences expressed in the human

XX  
CC brain. They can be used to measure gene expression in brain cell samples,

XX  
CC which may enable the diagnosis and improved treatment of nervous system

XX  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

XX  
CC epilepsy and cancers. The present sequence is one of the probes of the

XX  
CC invention.

XX  
SQ Sequence 552 BP; 163 A; 95 C; 201 G; 93 T; 0 other;

## RESULT 6

AAK37711

ID AAK37711 standard; DNA: 552 BP.

XX  
AC AAK37711;

XX  
DT 06-NOV-2001 (first entry)

XX  
DE Human bone marrow expressed single exon probe SEQ ID NO: 12268.

XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;

XX  
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.



CC The present invention relates to a Moraxella catarrhalis genomic library  
 CC comprising of a combination of 41 nucleic acid molecules (see  
 CC AAF28514-AAF28554). The library has a number of uses described in the  
 CC specification e.g. is useful for identifying diagnostic and therapeutic  
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large  
 CC aerobic, gram-negative diplococcus, normally found among the bacterial  
 CC flora of human upper airways. M. catarrhalis is known to cause acute,  
 CC localised infections such as otitis media, sinusitis and bronchopulmonary  
 CC infection and life-threatening, systemic diseases including endocarditis  
 CC and meningitis.  
 XX  
 SQ Sequence 89047 BP; 26501 A; 17338 C; 19060 G; 26147 T; 1 other;

Query Match 82.9%; Score 17.4; DB 22; Length 89047;  
 Best Local Similarity 94.7%; Pred. No. 1.1e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 3 catgggagatgttggcagc 21  
 ||| ||||| ||||| |||||  
 Db 38311 catcgagatgttggcagc 38329

RESULT 9  
 AAQ03742  
 ID AAQ03742 standard; DNA; 2839 BP.

AC AAQ03742;

DT 15-AUG-1989 (first entry)

DE Human SKI related gene.

DE Human sloan kettering institute related gene; cancer; ss.

DE Key Location/Qualifiers

DE CDS 709..2765

DE FT /\*tag= a

DE FT /label=human SKI-related protein

DE PN JP02053485-A.

DE PD 22-FEB-1990.

DE PF 19-AUG-1988; 88JP-0205906.

DE PR 19-AUG-1988; 88JP-0205906.

DE PA (RIKA ) RIKAGAKU KENKYUSHO .

DE WPI: 1990-103117-A.

DE P-PSDB: AAR03664.

DE Human SKI related gene - coded by base sequence including 709 adenine to

DE 1806 guanine etc.

DE PS Disclosure; fig 1; 6pp; Japanese.

DE This sequence encodes a human sloan kettering institute (SKI) related

DE protein. A truncated protein can also be encoded by bases 709-1806.

DE The protein can be used as a cancer marker by preparing an antibody

DE against it. The copy no. of a cancer gene can be found by Southern

DE blotting using the gene or fragments. The degree of malignancy of a

DE cancer can thus be determined. See also AAQ03743.

DE SQ Sequence 2839 BP; 936 A; 521 C; 586 G; 796 T; 0 other;

Query Match 80.0%; Score 16.8; DB 11; Length 2839;

Best Local Similarity 90.0%; Pred. No. 1.4e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 gcatgggagatgttggcagc 21

Db 767 ggatgggagatgttggcagc 786

RESULT 10

AAQ03743

ID AAQ03743 standard; DNA; 2888 BP.

AC AAQ03743;

DT 15-AUG-1989 (first entry)

DE Human SKI related gene.

DE Human sloan kettering institute related gene; cancer; ss.

DE Key Location/Qualifiers

DE CDS 709..1956

DE FT /\*tag= a

DE FT /label=human SKI related protein

DE PN JP02053485-A.

DE PD 22-FEB-1990.

DE PF 19-AUG-1988; 88JP-0205906.

DE PR 19-AUG-1988; 88JP-0205906.

DE PA (RIKA ) RIKAGAKU KENKYUSHO .

DE WPI: 1990-103117-A.

DE P-PSDB: AAR03665.

DE Human SKI related gene - coded by base sequence including 709 adenine to

DE 1806 guanine etc.

DE PS Disclosure; fig 2; 6pp; Japanese.

DE This sequence encodes a human sloan kettering institute (SKI) related

DE protein. The protein can be used as a cancer marker by preparing an

DE antibody against it. The copy no. of a cancer gene can be found by

DE Southern blotting using the gene or fragments. The degree of malignancy

DE of a cancer can thus be determined. See also AAQ03742.

DE SQ Sequence 2888 BP; 857 A; 566 C; 598 G; 867 T; 0 other;

Query Match 80.0%; Score 16.8; DB 11; Length 2888;

Best Local Similarity 90.0%; Pred. No. 1.4e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 gcatgggagatgttggcagc 21

Db 768 ggatgggagatgttggcagc 787

RESULT 11

AAS66199/c

ID AAS66199 standard; cDNA; 1026 BP.

AC AAS66199;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #2003.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;

DE food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
XX P-PSDB; ABG02012.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity -  
XX  
XX Claim 1; SEQ ID No 2003; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences, (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. AAS64197-AAS94564 represent novel human  
XX diagnostic coding sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 1026 BP; 245 A; 292 C; 289 G; 200 T; 0 other;  
XX  
Query Match 78.1%; Score 16.4; DB 23; Length 1026;  
Best Local Similarity 94.4%; Pred. No. 1.9e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
Qy 4 atggagagatgttgcagc 21  
Db 447 ATGGGAGAGTGTGGCATC 430  
XXXXXXXXXXXXXXXXXXXX  
RESULT 12  
AAH92223/C  
ID AAH92223 standard; DNA; 700 BP.  
XX  
XX AC AAH92223;  
XX  
XX DT 09-OCT-2001 (first entry)  
XX  
XX Human inflammatory bowel disease related gene fragment IGR2235a.  
XX  
XX Human; inflammatory bowel disease; Crohn's disease; ulcerative colitis;  
XX single nucleotide polymorphism; SNP; chromosome 19p13; paternity test;  
XX chromosome 5q31-33; forensic test; gene therapy; ds.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200142511-A2.

XX PD 14-JUN-2001.  
XX PF 11-DEC-2000; 2000WO-US33632.  
XX PR 10-DEC-1999; 99US-0170257.  
XX PR 10-APR-2000; 2000US-0196046.  
XX  
XX (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
XX PA (ELLI-) ELLIPSIS BIOTHERAPEUTICS CORP.  
XX  
XX Daly M, Hudson TJ, Lander ES, Rioux J, Siminovitch K;  
XX WPI; 2001-367874/38.  
XX  
XX Testing for the presence of polymorphisms associated with inflammatory  
XX bowel disease, using a hybridization assay -  
XX  
XX Disclosure; Page 154; 463pp; English.  
XX  
XX The present invention describes a method for detecting the presence of  
XX polymorphisms associated with inflammatory bowel diseases such as  
XX ulcerative colitis and Crohn's disease. The methods can be used to detect  
XX the presence of genetic polymorphisms associated with inflammatory bowel  
XX disease and correlating their occurrence with disease states. They may be  
XX used in this way for phenotypic correlations, forensics, paternity  
XX testing, medicine and genetic analysis. The present sequence is a gene  
XX containing a polymorphic site described in the exemplification of the  
XX invention.  
XX  
XX Sequence 700 BP; 191 A; 168 C; 163 G; 178 T; 0 other;  
XX  
Query Match 77.1%; Score 16.2; DB 22; Length 700;  
Best Local Similarity 85.7%; Pred. No. 2.3e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
XX  
Qy 1 agcatggagagatgttgcagc 21  
Db 667 AGCATGGGAGAGTGTGACATC 647  
XXXXXXXXXXXXXXXXXXXX  
RESULT 13  
AAH92224/C  
ID AAH92224 standard; DNA; 700 BP.  
XX  
XX AC AAH92224;  
XX  
XX DT 09-OCT-2001 (first entry)  
XX  
XX Human inflammatory bowel disease related gene fragment IGR2236a.  
XX  
XX Human; inflammatory bowel disease; Crohn's disease; ulcerative colitis;  
XX single nucleotide polymorphism; SNP; chromosome 19p13; paternity test;  
XX chromosome 5q31-33; forensic test; gene therapy; ds.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200142511-A2.  
XX  
XX PD 14-JUN-2001.  
XX  
XX PF 11-DEC-2000; 2000WO-US33632.  
XX PR 10-DEC-1999; 99US-0170257.  
XX PR 10-APR-2000; 2000US-0196046.  
XX  
XX (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
XX PA (ELLI-) ELLIPSIS BIOTHERAPEUTICS CORP.  
XX  
XX Daly M, Hudson TJ, Lander ES, Rioux J, Siminovitch K;  
XX WPI; 2001-367874/38.

XX Testing for the presence of polymorphisms associated with inflammatory  
PT bowel disease, using a hybridization assay -  
XX  
XX  
PS Disclosure; Page 154; 463pp; English.  
XX  
CC The present invention describes a method for detecting the presence of  
CC polymorphisms associated with inflammatory bowel diseases such as  
CC ulcerative colitis and Crohn's disease. The methods can be used to detect  
CC the presence of genetic polymorphisms associated with inflammatory bowel  
CC disease and correlating their occurrence with disease states. They may be  
CC used in this way for phenotypic correlations, forensics, paternity  
CC testing, medicine and genetic analysis. The present sequence is a gene  
CC containing a polymorphic site described in the exemplification of the  
CC invention.  
XX  
SQ Sequence 700 BP; 197 A; 178 C; 160 G; 165 T; 0 other;

Query Match 77.1%; Score 16.2; DB 22; Length 700;  
Best Local Similarity 85.7%; Pred. No. 2.3e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 agcatggagatgttgcgcac 21  
||||||| |||||  
Db 168 AGCATGGAGAGGTTGCACATC 148

RESULT 14  
AAC46188/c  
ID AAC46188 standard; DNA; 935 BP.  
XX  
AC AAC46188;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 49219.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 03-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 99US-0134256.

14-MAY-1999; 99US-0134218.  
14-MAY-1999; 99US-0134219.  
14-MAY-1999; 99US-0134221.  
14-MAY-1999; 99US-0134370.  
18-MAY-1999; 99US-0134768.  
19-MAY-1999; 99US-0134941.  
20-MAY-1999; 99US-0135124.  
21-MAY-1999; 99US-0135353.  
24-MAY-1999; 99US-0135629.  
25-MAY-1999; 99US-0136021.  
27-MAY-1999; 99US-0136392.  
28-MAY-1999; 99US-0136782.  
01-JUN-1999; 99US-0137222.  
03-JUN-1999; 99US-0137528.  
04-JUN-1999; 99US-0137502.  
07-JUN-1999; 99US-0137724.  
08-JUN-1999; 99US-0138094.  
10-JUN-1999; 99US-0138540.  
10-JUN-1999; 99US-0138847.  
14-JUN-1999; 99US-0139119.  
16-JUN-1999; 99US-0139452.  
16-JUN-1999; 99US-0139453.  
17-JUN-1999; 99US-0139492.  
18-JUN-1999; 99US-0139454.  
18-JUN-1999; 99US-0139455.  
18-JUN-1999; 99US-0139456.  
18-JUN-1999; 99US-0139462.  
18-JUN-1999; 99US-0139463.  
18-JUN-1999; 99US-0139750.  
18-JUN-1999; 99US-0139763.  
21-JUN-1999; 99US-0139817.  
22-JUN-1999; 99US-0139899.  
23-JUN-1999; 99US-0140353.  
23-JUN-1999; 99US-0140354.  
24-JUN-1999; 99US-0140695.  
28-JUN-1999; 99US-0140823.  
29-JUN-1999; 99US-0140991.  
30-JUN-1999; 99US-0141287.  
01-JUL-1999; 99US-0141842.  
01-JUL-1999; 99US-0142154.  
02-JUL-1999; 99US-0142055.  
06-JUL-1999; 99US-0142390.  
08-JUL-1999; 99US-0142803.  
09-JUL-1999; 99US-0142920.  
12-JUL-1999; 99US-0142977.  
13-JUL-1999; 99US-0143542.  
14-JUL-1999; 99US-0143624.  
15-JUL-1999; 99US-0144005.  
16-JUL-1999; 99US-0144085.  
16-JUL-1999; 99US-0144086.  
19-JUL-1999; 99US-0144325.  
19-JUL-1999; 99US-0144331.  
19-JUL-1999; 99US-0144332.  
19-JUL-1999; 99US-0144333.  
19-JUL-1999; 99US-0144334.  
19-JUL-1999; 99US-0144335.  
20-JUL-1999; 99US-0144352.  
20-JUL-1999; 99US-0144632.  
20-JUL-1999; 99US-0144884.  
21-JUL-1999; 99US-0144814.  
21-JUL-1999; 99US-0145086.  
21-JUL-1999; 99US-0145088.  
22-JUL-1999; 99US-0145085.  
22-JUL-1999; 99US-0145087.  
22-JUL-1999; 99US-0145089.  
22-JUL-1999; 99US-0145192.  
23-JUL-1999; 99US-0145145.  
23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 26-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.

PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 77.1%; Score 16.2; DB 21; Length 935;  
Best Local Similarity 85.7%; Pred. No. 2.4e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 agcatggagagatgttggcagc 21  
||||| ||||| ||||| ||  
DB 670 AGCATGGGAGACGTGGCTGC 650

RESULT 15  
AAC35293/c  
ID AAC35293 standard; DNA; 938 BP.  
XX  
AC AAC35293;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 9662.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140595.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143442.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 02-AUG-1999; 99US-0145951.  
PR 99US-0146386.  
PR 02-AUG-1999; 99US-0146387.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.



PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 77.1%; Score 16.2; DB 21; Length 938;  
Best Local Similarity 85.7%; Pred. No. 2.4e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 agcatgggagatgttggcagc 21  
|||||  
Db 670 ACCATGGGAGACGGTGCTGC 650

Search completed: May 18, 2002, 17:46:03  
Job time: 13752 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 14:59:01 ; Search time 2878.96 Seconds  
(without alignments)  
152.644 Million cell updates/sec

Title: US-09-623-329-25

Perfect score: 21

Sequence: 1 agcatggagatgttgcagc 21

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.pro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htgo.inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----					

RESULT	1	AX018466	Sequence 25 from Patent WO9945155.	21 bp	DNA	linear	PAT 07-SEP-2000
LOCUS	AX018466	Sequence 25 from Patent WO9945155.					
DEFINITION	AX018466	Sequence 25 from Patent WO9945155.					
ACCESSION	AX018466	Sequence 25 from Patent WO9945155.					
VERSION	AX018466.1	GI:10042617					
KEYWORDS		Epstein-Barr virus.					
SOURCE		Human herpesvirus 4					
ORGANISM		Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae; Lymphocryptovirus.					
REFERENCE		1 (bases 1 to 21)					
AUTHORS		Middelorp J.M., Van Den Brule A.J. and Vervoot M.B.					
TITLE		Oligonucleotides for the amplification and detection of Epstein Barr virus (ebv) nucleic acid					
JOURNAL		Patent: WO 9945155-A 25 10-SEP-1999;					
FEATURES		MIDDELORP JAAP MICHEL (NL); AKZO NOBEL NV (NL); DEN BRULE ADRIANUS JOHANNES CH (NL); VERVOORT MARCEL BARTOLINA HEND (NL); Location/Qualifiers					
source		1. .21					
		/organism="Human herpesvirus 4"					
		/db_xref="taxon:10376"					
BASE COUNT		5 a 3 c 9 g 4 t					

## ALIGNMENTS

AX018466 Sequence  
V01555 Epstein-Bar  
M80517 Epstein-Bar  
AX313760 Sequence  
U02619 Human TFI1  
AC087076 Caenorhab  
AL138923 Human DNA  
AC023303 Homo sapi  
AC026691 Homo sapi  
AC025125 Homo sapi  
AC027040 Homo sapi  
AC027796 Homo sapi  
AC084737 Homo sapi  
AC099564 Homo sapi  
AC019149 Homo sapi  
AC015840 Homo sapi  
AP003127 Homo sapi  
AF222337 Rattus no  
AF222338 Rattus no  
M92905 Rat calcium  
AF055477 Rattus no  
AC016257 Homo sapi  
AC078874 Homo sapi  
AC099233 Rattus no  
AC104805 Homo sapi  
AC004673 Homo sapi  
AF096286 Mus muscu  
AL135782 Homo sapi  
AX067459 Sequence  
Continuation (2 of  
AC025059 Homo sapi  
AC103298 Rattus no  
AC004624 Homo sapi  
AC005289 Homo sapi  
AC070017 Giardia i  
X91923 H.sapiens E  
AJ011770 Homo sapi  
AT251580 Drosophila  
U070730 Human SNO2  
E02251 human 'sno  
X15219 Human sno o  
X15217 Human sno o  
E02252 human 'sno  
Z19588 H.sapiens s



REFERENCE 18 (bases 1 to 172281)  
 AUTHORS Farrell,P.J. and Barrell,B.G.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-JUN-1984)  
 REFERENCE 19 (bases 1 to 172281)  
 AUTHORS Farrell,P.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-MAR-1988) Farrell P., Ludwig Institute for Cancer Research, St. Mary's Hospital Medical School, Norfolk Place London W2 1PG

COMMENT

Listed under this feature are all known protein coding regions as well as all the major open reading frames in the sequence. In general the term major is taken as the longest frame in a particular region taking into account the adjacent longest frames and likely transcription signals. Note that on this basis some long overlapping frames have been excluded and on the other hand some small frames have been included which might represent exons or genes because they occur in a logical combination with other features or because of some other experimental data. The reading frames are named according to the Bam HI fragment in which they start. eg BALE3 is the third leftward frame starting in Bam HI fragment A. BOP1 is the first rightward frame in Bam HI fragment O. If there is an obvious TATA sequence followed by an in frame Met codon that satisfies the rules of Kozak [12] in that there is a purine at -3 and/or a G at +4 then the reading frame is numbered from the A of the ATG to the base preceding the termination codon. If there is no obvious initiation codon or there is a substantial reading frame in phase before the ATG then the reading frame is numbered from the first base of the first codon.

SITES OF POLYA signals  
 This feature lists all occurrences of the sequence AATAAA which is found normally approximately 20 bases upstream of the mRNA processing/polyA addition site. The rarely used homolog AATAAA is only listed when it is found in a position close to the end of a major reading frame.

SITES OF DONOR and ACCEPT sequences  
 This is not a comprehensive listing of all such sequences and only the positions of a few have been noted because they occur in potentially interesting positions. The number quoted in the table is the position of the terminal base in the intron in each case. Restriction enzyme SITES.  
 Only the positions of the sites Bam HI (BAM) are listed.

RPT  
 This feature is used to define repetitive sequences.

SITE DEL  
 This feature defines deletions in B95-8 with respect to other strains such as RAJI and also to deletions in other strains such as P3HR1 and DAUDI with respect to B95-8.

SITE HPN  
 Denotes sequences with twofold symmetry ie could form hairpin loops. This is not a comprehensive list - only a few occurrences noted.

ORGRPL  
 Denotes the region that encompasses an origin of replication (ori P). [13].

NUMBERING  
 The DNA sequence of B95-8 EBV has been revised [19]. The original (Baer et al, 1984) base 359 has been deleted so the new sequence around that position reads TCAGTCTTT. To avoid renumbering the entire sequence, position 1 has been moved 1 base to the left of the EcoRI site separating EcoRI Dhet from EcoRI 1 (ie the first A of AGAATTC).

FEATURES

source  
 1. 172281  
 /organism="Human herpesvirus 4"  
 /strain="B95-8"  
 /db\_xref="taxon:10376"

Query Match 90.5%; Score 19; DB 14; Length 172281;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 catggggagatgttggcagc 21  
 |||||  
 Db 165798 CATCGGAGATGTTGGCAGC 165780

RESULT 3

HS4B958RAJ/c  
 LOCUS 184113 bp DNA linear VRL 12-APR-1996  
 DEFINITION Epstein-Barr virus, artifactual joining of B95-8 complete genome and the sequences from Raji of the large deletion found in B95-8.

ACCESSION M80517 M75989  
 VERSION M80517.1 GI:330330  
 KEYWORDS Human herpesvirus 4 DNA.  
 SOURCE Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae; Lymphocryptovirus.

ORGANISM

REFERENCE 1 (sites)  
 AUTHORS Baer,R.J., Bankier,A.T., Biggin,M.D., Deininger,P.L., Farrell,P.J., Gibson,T.J., Hatfull,G.F., Hudson,G.S., Satchwell,S.C., Seguin,C., Tuffnell,P.S. and Barrell,B.G.

TITLE

JOURNAL Nature 310 (5974), 207-211 (1984)  
 MEDLINE 84270667  
 REFERENCE 2 (sites)  
 AUTHORS Parker,B.D., Bankier,A., Satchwell,S., Barrell,B. and Farrell,P.J.

TITLE

JOURNAL Virology 179 (1), 339-346 (1990)  
 MEDLINE 91021036  
 REFERENCE 3 (sites)  
 AUTHORS Sample,J., Brooks,L., Sample,C., Young,L., Rowe,M., Gregory,C., Rickinson,A. and Kieff,E.

TITLE

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 88 (14), 6343-6347 (1991)  
 MEDLINE 91296817  
 REFERENCE 4 (bases 1 to 184113)  
 AUTHORS Jenson,H.B.

TITLE

JOURNAL Unpublished (1992)  
 COMMENT The B95-8 genome (V01555) has a large deletion in the right side of the genome which has been sequenced in Raji (M35547). These sequences have been joined to form an extended and more complete, although artifactual, EBV sequence.

FEATURES

source  
 1. 184113  
 /organism="Human herpesvirus 4"  
 /db\_xref="taxon:10376"  
 misc\_feature  
 1. 152008  
 /note="B95-8 sequences (corresponds to 1-152,008 of V01555)"  
 misc\_feature  
 152009..152012  
 /note="Overlap of B95-8 and Raji sequences at B95-8 deletion point (corresponds to 152,009-152,012 in V01555, and 1-4 in M35547)"  
 misc\_feature  
 153013..163839  
 /note="Raji sequences (corresponds to 5-11,831 of M35547)"  
 misc\_feature  
 163840..163843  
 /note="overlap of B95-8 and Raji sequences at B95-8 deletion point (corresponds to 152,009-152,012 of V01555, and 11,832-11,835 of M35547)"  
 misc\_feature  
 163844..184113  
 /note="B95-8 sequences (corresponds to 152,013-172,282 of V01555)"

BASE COUNT 36002 a 55824 c 54622 g 37665 t  
 ORIGIN

Query Match 90.5%; Score 19; DB 14; Length 184113;  
 Best Local Similarity 100.0%; Pred. No. 44;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 catggagatgttgcagc 21  
|||||

Db 177629 CATGGGAGATGTTGGCAGC 177611

## RESULT 4

AX313760/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

19; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

QY 2 gcatggagatgttgcagc 21

|||||

Db 172 GCATGGTAGATGTTGGCAGC 153

## RESULT 5

HSU02619/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

19; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

QY 2 gcatggagatgttgcagc 21

|||||

Db 172 GCATGGTAGATGTTGGCAGC 153

|||||

Db 172 GCATGGTAGATGTTGGCAGC 153

|||||

Db 172 GCATGGTAGATGTTGGCAGC 153

|||||

/codon\_start=1  
/product="TFIIIC Box B-binding subunit"  
/protein\_id="AA117985.1"  
/db\_xref="GI:442362"  
/translation="MDALESILDEVALEGLDGLCLPALWSRLTRVPPFLPIEPCPTQ  
EFLWRALATHPGISFYEEPREKPDLOQDRYEEIDLEIGLESKRDPVALEDDVPIHM  
ILENKDGIOSCRYFKERKNTNDITRTKSLQPCMTVEPEDFMGKLLIGSLPAHAYO  
ALDSPGGGSRPEARALLLHPGTARPVOGELQRLDHTTAFKVDAGKLYHKLILNK  
NLITMOSHVIRLPTGAQHSILILNRFHVRKSYDILMEKLSVMISTRNNHIEL  
GKLEELGCEFTKRLQYIMINAGLAKVSLRQEIHEPCGPTKTKGDTGVMMVRLK  
LLKFKKNDHDDDEEVIKVTVPFMEDEGRQRTKYISCFVAFESDLRSQYQREKARS  
MNVGKLEARMCLRLORFKVKKPMPDEGQRTKYISCFVAFESDLRSQYQREKARS  
ELLTVSILASMOEESLLPEGHUFISESDEEERSSKRRGSGUSKDTASANLRPKT  
OPHSTPTKGGKVVNLPLKPPSPFGAERACUSLASRUSLDTSSVSPNVSF  
VSHCANSNGDIAVIEVRMENPKESSSLKTGRHSGQDKPIHTYLLKRNLLIEA  
VTNLRLESFTTQKIMIDQEGVCKCKKSTVRLVKNISSEGLRLRYRTVIVQ  
GJKKKVDIVVHPSMDQNDPLVSAIEQVRFRISSNSTANKVTSQPPVPOGEAEDSO  
KGEPSGDSQSASASSRSGSKMKSDNKGITPLRNYHPVIVPGLSRSLGFLPKMP  
LRVHMFELWYLIYHPASNTVEKPSFISERRTIKQESGRAGVPSGSSGSAWEACSEA  
PSKSGODGVTAWEAELATETVYVDDASMMRYIPPIVHRDFGFGWALVSDILCLPL  
SIPQIVOVSYKYVDNLEEFNDPLKHTLIRFLPRPKOQLLYKRRYIESVVENLQRL  
CYMGVLOFGPTKFDQKQDFILKKNAYIVDTTICDPHYNLGKRRPFRERRLYVLS  
MQDVNWFYDLQCVCLNTPLGVHPCVRKNSSTDQSGDEESLQKQESAMXKHNE  
RKAMLEYTTGSEVVDGLPGDGLGAAGLDSFYGHKRNMTWTSYIINQAKKENT  
AAENGLTVRLQTLFSLKRPMLSAKNSRLNIWGEARYGSELCAWEDQSAHHRMTRLR  
RNRVRGKSKOKRKLKDPCKIKRKGKGFEGEKSKRLRYHDEADQSAHHRMTRLR  
VTMSQEDGLLVLCRIASNVLTAKVKGPFVTVVVRDILHATFEESLDKTSKLGRRR  
RYIVKQAVLNYKVCIAENVYQDKALVGFDMNRKGYDDPKVCANEFKEVELKEKF  
SSALRNSLEIPDLOELFARYVLAIGDEKDOTRKEDELSVDDIHLVLQNLQST  
LALSQMSKTSQSYRTFLRYKDHVLRKAFMECQRKSLNRRKRVNHTLQPKKNRRL  
PFVPMYSQTSQYRTFLRYKDHVLRKAFMECQRKSLNRRKRVNHTLQPKKNRRL  
TNDMAFSLDGGPCNVAVLTLSGLISVDVRIPEQIIVDSWNEVYIKSLGKDG  
SLEHDEDEDDLEGCVGKRRMEVPAQASHNYLLMRGYISPGIVSTRLNPNDSI  
VNSCOMKFLQCTVPVRLPAAALEETMTGTSCLPTDTFTKLINPOENTCSLEEFV  
LOLESQVSPEDTLAELIATATCGDIDKEELRRRFALEKAGQGRTRTFADCI  
QALLEQVLEVCGNTARLVAMGSANPHLLSHVRLKDRDADTQREDPOARLEGSS  
EDSPGGAAPSPSPGRTKRASWASNETDAEGTONTAKRPAQDSNLAISLPGK  
AEDGAIAQAPSPPALEDAAGAAQDEQEGVEFSSPGQEQISGQAGPEGSDPGRG  
FTESFGAANIQAARERDESVCVIGRPVRVVDGHLNLVPCYKGMMEALYHIMTRPGI  
PESILRHVQGVLPVAVLELIQLESLSGCIKKRLRKRKPKVPVSLFTVPVVEVPSS  
LDESIMAFYEPTLDCTLRGVRFPHEVNNKNIHL"

BASE COUNT 1663 a 1959 c 2030 g 1444 t

ORIGIN

Query Match 87.6%; Score 18.4; DB 9; Length 6996;

Best Local Similarity 95.0%; Pred. No. 96;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcatggagatgttgcagc 21

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

|||||

Db 5972 GCCTGGAGATGTTGGCAGC 5953

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
Submitted by:

Genome Sequencing Center  
Department of Genetics, Washington University,  
St. Louis, MO 63110, USA  
e-mail: jspliehwatson.wustl.edu

COMMENT

NOTICE: This sequence may not be the entire insert of this clone.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we provide a small overlap between  
neighboring submissions.

FEATURES

source

1..83495  
/organism="Caenorhabditis briggsae"  
/strain="Gujarat G16"  
/db\_xref="taxon:6238"  
/clone="CB032N22"

BASE COUNT 24900 a 15987 c 15979 g 26629 t

Query Match 87.6%; Score 18.4; DB 3; Length 83495;  
Best Local Similarity 95.0%; Pred. No. 91;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agcatggagatgttggcag 20  
|||||

DB 43959 AGCATGGAGATGTGGCGG 43978

RESULT 7

AL138923/c

LOCUS

DEFINITION

Human DNA sequence from clone RP11-397J18 on chromosome 10,

ACCESSION

AL138923

VERSION

AL138923.9

KEYWORDS

HTG.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 85322)

REFERENCE

AUTHORS

Wray, P.

TITLE

JOURNAL

COMMENT

Submitted (17-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Aug 20, 2001 this sequence version replaced gi:15020724.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit submissions with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em: EMBL; Sw:  
SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 10, constructed by the Sanger Centre Chromosome 10  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr10  
RP11-397J18 is from the library RPCI-11.2 constructed by the group

of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone  
RP11-397J18. It may be shorter because we sequence overlapping  
sections only once, except for a 100 base overlap.

The true left end of clone RP11-429H9 is at 83323 in this sequence.  
The true right end of clone RP11-512J3 is at 2000 in this sequence.

FEATURES

source

Location/Qualifiers  
1..85322  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP11-397J18"  
/clone\_lib="RPCI-11.2"

BASE COUNT 25578 a 18206 c 17225 g 24311 t

ORIGIN

Query Match 87.6%; Score 18.4; DB 9; Length 85322;  
Best Local Similarity 95.0%; Pred. No. 91;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 agcatggagatgttggcag 20  
|||||

DB 34898 AGCTGGGAGATGTGGCAG 34879

RESULT 8

HUAC002303

LOCUS

DEFINITION

Homo sapiens Chromosome 16 BAC clone CIT987-SKA-670B5 -complete

ACCESSION

AC002303

VERSION

AC002303.1

KEYWORDS

HTG.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 131530)

REFERENCE

AUTHORS

Adams, M.D., Loftus, B.J., Zhou, L., LaBombard, M., Fuhrmann, J.,

Brandon, R., Kim, U.J., Kerlavage, A.R. and Venter, J.C.

TITLE

JOURNAL

MEDLINE

PURMED

REFERENCE

AUTHORS

Adams, M.D., Loftus, B.J., Zhou, L., LaBombard, M., Fuhrmann, J.,

Brandon, R., Kim, U.J., Kerlavage, A.R. and Venter, J.C.

TITLE

JOURNAL

REFERENCE

AUTHORS

Adams, M.D.

TITLE

JOURNAL

COMMENT

Submitted (25-AUG-1997) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA  
On Aug 26, 1997 this sequence replaced gi:2317824.  
Address all correspondence to:  
Mark Adams  
The Institute for Genomic Research  
9712 Medical Center Dr.,  
Rockville, MD 20850,







```

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence-submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7767
Center clone name: 147_K_16
-----
FEATURES             Location/Qualifiers
     source            1..157481
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="RP11-147K16"
                        /clone_lib="RPC1-11 Human Male BAC"
     repeat_region    complement(49..183)
                        /rpt_family="MIR3"
     repeat_region    560..671
                        /rpt_family="MIR"
     repeat_region    complement(1322..1611)
                        /rpt_family="AluSx"
     repeat_region    complement(1971..2254)
                        /rpt_family="AluSx"
     repeat_region    2257..2321
                        /rpt_family="MIR3"
     repeat_region    /rpt_family="MIR"
     repeat_region    2359..2484
                        /rpt_family="MER4C"
     repeat_region    complement(2485..2784)
                        /rpt_family="AluSx"
     repeat_region    2785..2891
                        /rpt_family="MER4C"
     repeat_region    /rpt_family="MER4C"
     repeat_region    complement(2934..3246)
                        /rpt_family="AluSx"
     repeat_region    3266..3458
                        /rpt_family="MER4C"
     repeat_region    3794..3838
                        /rpt_family="(CCCCAG)n"
     repeat_region    complement(3913..4235)
                        /rpt_family="AluSx"
     repeat_region    4645..4951
                        /rpt_family="AluSg"
     repeat_region    complement(5564..5677)
                        /rpt_family="L2"
     repeat_region    complement(6590..6691)
                        /rpt_family="L2"
     repeat_region    6694..7004
                        /rpt_family="AluSx"
     repeat_region    complement(7564..7866)
                        /rpt_family="AluSx"
     repeat_region    8283..8405
                        /rpt_family="L2"
     repeat_region    complement(8851..9002)
                        /rpt_family="AluSx"
     repeat_region    9003..9078
                        /rpt_family="MIR"
     repeat_region    9168..9190
                        /rpt_family="AT-rich"
     repeat_region    complement(9254..9578)
                        /rpt_family="AluJb"
     repeat_region    complement(9735..10065)
                        /rpt_family="L1MC4a"
     repeat_region    10066..10358
                        /rpt_family="AluSx"
     repeat_region    complement(10359..10724)
                        /rpt_family="L1MC4a"
     repeat_region    complement(10744..11032)
                        /rpt_family="AluY"
     repeat_region    complement(11036..11328)
                        /rpt_family="AluSx"
     repeat_region    11426..11546
                        /rpt_family="AluJo"

```

```

repeat_region    complement(11556..11684)
                        /rpt_family="L1ME4A"
repeat_region    11691..11855
                        /rpt_family="AluJo"
repeat_region    complement(12336..12623)
                        /rpt_family="AluSg"
repeat_region    12759..13059
                        /rpt_family="L2"
repeat_region    13409..13450
                        /rpt_family="G-rich"
repeat_region    13451..13792
                        /rpt_family="GA-rich"
                        /note="<30 qual SNGL region"
unsure            13793..13822
                        /rpt_family="polypurine"
repeat_region    complement(14075..14128)
                        /rpt_family="MIR"
repeat_region    complement(14533..14679)
                        /rpt_family="AluSg/X"
repeat_region    complement(14943..15153)
                        /rpt_family="L2"
                        15154..15453
                        /rpt_family="AluSc"
repeat_region    complement(15454..15856)
                        /rpt_family="L2"
repeat_region    complement(17742..17803)
                        /rpt_family="MER5B"
repeat_region    complement(17864..18176)
                        /rpt_family="AluSg"
repeat_region    complement(18177..18239)
                        /rpt_family="MER5B"
repeat_region    complement(19538..19711)
                        /rpt_family="MIR3"
repeat_region    20990..21075
                        /rpt_family="MIR"
repeat_region    21095..21219
                        /rpt_family="(CCCCAG)n"
repeat_region    21232..21365
                        /rpt_family="(TCCA)n"
repeat_region    complement(21366..21764)
                        /rpt_family="L2"
repeat_region    22307..22357
                        /rpt_family="Alu"
repeat_region    complement(22450..22753)
                        /rpt_family="AluSg"
repeat_region    23036..23328
                        /rpt_family="AluSx"
repeat_region    complement(23329..23607)
                        /rpt_family="L1MH2"
repeat_region    23632..23912
                        /rpt_family="AluSx"
repeat_region    complement(24128..24260)
                        /rpt_family="ELAM_C"
repeat_region    complement(24797..24853)
                        /rpt_family="MIR"
repeat_region    24992..25025
                        /rpt_family="MER63C"
repeat_region    25026..25056
                        /rpt_family="(TG)n"
repeat_region    25057..25206

```

Query Match 87.6%; Score 18.4; DB 9; Length 157481;  
 Best Local Similarity 95.0%; Pred. No. 90;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcattggagatgttgccagc 21  
 |||||  
 Db 5207 GCATGGAGATGTTGGCGGC 5226

RESULT 11  
 AC027040

```

LOCUS       AC027040               159397 bp    DNA    linear    HTG 20-APR-2000
DEFINITION  Homo sapiens chromosome 6 clone RP11-601L9 map 6, WORKING DRAFT
ACCESSION   AC027040
VERSION     AC027040.2 GI:7596822
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 159397)
AUTHORS     Birren,B., Linton,L., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
             Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bida,F.,
             Boguslavskiy,L., Boukhalter,B., Brown,A., Burkett,G.,
             Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
             Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
             Galagan,J., Gardyna,S., Glinde,S., Goyette,M., Graham,L.,
             Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
             Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
             Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
             Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
             McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
             Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
             Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
             O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
             Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
             Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
             Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
             Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigillo,J.,
             Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
             Young,G., Zainoun,J., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (26-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 19, 2000 this sequence version replaced gi:7329401.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project information
Center project name: L6509
Center clone name: 601_L_9
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 145545 bases at least Q40
Consensus quality: 152293 bases at least Q30
Consensus quality: 155026 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 156997; sum-of-contigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 4.2 in Q20 bases; sum-of-contigs
-----
* NOTE: this is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1036: contig of 1036 bp in length
* 1037 1136: gap of 100 bp

```

```

* 1137 2280: contig of 1144 bp in length
* 2281 2380: gap of 100 bp
* 2381 3832: contig of 1452 bp in length
* 3833 3932: gap of 100 bp
* 3933 6727: contig of 2795 bp in length
* 6728 6827: gap of 100 bp
* 6828 10325: contig of 3498 bp in length
* 10326 10425: gap of 100 bp
* 10426 12401: contig of 1976 bp in length
* 12402 12501: gap of 100 bp
* 12502 15301: contig of 2800 bp in length
* 15302 15401: gap of 100 bp
* 15402 19319: contig of 3918 bp in length
* 19320 19419: gap of 100 bp
* 19420 22845: contig of 3426 bp in length
* 22846 22945: gap of 100 bp
* 22946 27420: contig of 4475 bp in length
* 27421 27520: gap of 100 bp
* 27521 31950: contig of 4430 bp in length
* 31951 32050: gap of 100 bp
* 32051 37858: contig of 5808 bp in length
* 37859 37958: gap of 100 bp
* 37959 41562: contig of 3604 bp in length
* 41563 41662: gap of 100 bp
* 41663 46390: contig of 4728 bp in length
* 46391 46490: gap of 100 bp
* 46491 51794: contig of 5304 bp in length
* 51795 51894: gap of 100 bp
* 51895 60772: contig of 8878 bp in length
* 60773 60872: gap of 100 bp
* 60873 68644: contig of 7772 bp in length
* 68645 68744: gap of 100 bp
* 68745 76352: contig of 7608 bp in length
* 76353 76452: gap of 100 bp
* 76453 83748: contig of 7296 bp in length
* 83749 83848: gap of 100 bp
* 83849 90391: contig of 6543 bp in length
* 90392 90491: gap of 100 bp
* 90492 99114: contig of 8623 bp in length
* 99115 99214: gap of 100 bp
* 99215 108728: contig of 9514 bp in length
* 108729 108828: gap of 100 bp
* 108829 116454: contig of 7626 bp in length
* 116455 116554: gap of 100 bp
* 116555 131109: contig of 14555 bp in length
* 131110 131209: gap of 100 bp
* 131210 159397: contig of 28188 bp in length.

```

```

FEATURES             Location/Qualifiers
     source            1..159397
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6"
     clone="RP11-601L9"
     clone_lib="RPC1-11 Human Male BAC"
     misc_feature      1..1036
                     /note="assembly_fragment"
                     1137..2280
     misc_feature      /note="assembly_fragment"
                     2381..3832
     misc_feature      /note="assembly_fragment"
                     3933..6727
     misc_feature      /note="assembly_fragment"
                     6828..10325
     misc_feature      /note="assembly_fragment"
                     10426..12401
     misc_feature      /note="assembly_fragment"
                     12502..15301
     misc_feature      /note="assembly_fragment"
                     15402..19319
     misc_feature      /note="assembly_fragment"
                     19420..22845

```

```

misc_feature 22946..27420
/note="assembly_fragment"
misc_feature 27521..31950
/note="assembly_fragment"
misc_feature 32051..37858
/note="assembly_fragment"
misc_feature 37959..41562
/note="assembly_fragment"
misc_feature 41663..46390
/note="assembly_fragment"
misc_feature 46491..51794
/note="assembly_fragment"
misc_feature 51895..60772
/note="assembly_fragment"
misc_feature 60873..68644
/note="assembly_fragment"
misc_feature 68745..76352
/note="assembly_fragment"
misc_feature 76453..83748
/note="assembly_fragment"
misc_feature 83849..90391
/note="assembly_fragment"
misc_feature 90492..99114
/note="assembly_fragment"
misc_feature 99215..108728
/note="assembly_fragment"
misc_feature 108829..116454
/note="assembly_fragment"
misc_feature 116555..131109
/note="assembly_fragment"
misc_feature 131210..159397
/note="assembly_fragment"
BASE COUNT 35835 a 41406 c 43302 g 36437 t 2417 others
ORIGIN

Query Match 87.6%; Score 18.4; DB 2; Length 159397;
Best Local Similarity 95.0%; Pred. No. 90;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcatggagatgttgcgcg 21
|||||
DB 33473 GCATGGAGATGTGCGGC 33492

RESULT 12
AC027796 163215 bp DNA linear HTG 27-JAN-2002
LOCUS Homo sapiens chromosome 17 clone RP11-235E17 map 17, *** SEQUENCING
DEFINITION IN PROGRESS ***, 4 unordered pieces.
ACCESSION AC027796
VERSION AC027796.6 GI:18001687
KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Birren.B., Linton.L., Nusbaum.C. and Lander.E.
TITLE Homo sapiens chromosome 17, clone RP11-235E17
JOURNAL Unpublished
REFERENCE
AUTHORS Birren.B., Linton.L., Nusbaum.C., Lander.E., Abraham.H., Allen.N.,
Anderson.S., Baldwin.J., Barna.N., Bastien.V., Beda.F.,
Boguslavsky.L., Bouckgalter.B., Brown.A., Burkett.G.,
Campotiano.A., Castle.A., Choepel.Y., Colangelo.M., Collins.S.,
Collimore.A., Cooke.P., Dearellano.K., Dewar.K., Diaz.J.S.,
Dodge.S., Domino.M., Doyle.M., Ferreira.P., FitzHugh.W., Gage.D.,
Galagan.J., Gardyna.S., Ginde.S., Goyette.M., Graham.L.,
Grand-Pierre.N., Grant.G., Hagos.B., Heaford.A., Horton.L.,
Howland.J.C., Iliev.I., Johnson.R., Jones.C., Kann.L., Karatas.A.,
Klein.J., LaRoque.K., Lamazares.R., Landers.T., Lechoczky.J.,
Levine.R., Lieu.C., Liu.G., Locke.K., Macdonald.P., Marquis.N.,
McCarthy.M., McEwan.P., McGurk.A., McKernan.K., McPheeters.R.,
Meldrim.J., Meneus.L., Mihova.T., Miranda.C., Mlenka.V., Morrow.J.,
Murphy.T., Naylor.J., Norman.C.H., O'Connor.T., O'Donnell.P.,
O'Neill.D., Oliver.T.M., Oliver.J., Peterson.K., Pierre.N.,
Pisani.C., Pollara.V., Raymond.C., Riley.R., Rogov.P., Rothman.D.,
Roy.A., Santos.R., Schauer.S., Severy.P., Spencer.B.,
Stange-Thomann.N., Stojanovic.N., Subramanian.A., Talamas.J.,
Tefaye.S., Theodore.J., Tirrell.A., Travers.M., Trigilio.J.,
Vassiliev.H., Viel.R., Vo.A., Wilson.B., Wu.X., Wyman.D., Ye.W.J.,
Young.G., Zainoun.J., Zimmer.A. and Zody.M.
Direct Submission
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 31, 2001 this sequence version replaced gi:14030095.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBK
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9035
Center clone name: 235_E_17
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 63927: contig of 63927 bp in length
* 3928 64027: gap of 100 bp
* 4028 127595: contig of 63568 bp in length
* 127596 127695: gap of 100 bp
* 127696 153978: contig of 26283 bp in length
* 153979 154078: gap of 100 bp
* 154079 163215: contig of 9137 bp in length.
Location/Qualifiers
1..163215
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-235E17"
/clone_lib="RPC1-11 Human Male BAC"
BASE COUNT 37312 a 41231 c 42820 g 41370 t 482 others
ORIGIN

Query Match 87.6%; Score 18.4; DB 2; Length 163215;
Best Local Similarity 95.0%; Pred. No. 90;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 gcatggagatgttgcgcg 21
|||||
DB 119309 GCATGGAGATGTGCGGC 119328

RESULT 13
AC084737/c 166458 bp DNA linear HTG 11-NOV-2000
LOCUS Homo sapiens chromosome UL clone RP13-516M14, WORKING DRAFT
DEFINITION SEQUENCE, 25 unordered pieces.
ACCESSION AC084737
VERSION AC084737.2 GI:11138191
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 166458)  
Waterston, R.H.  
The sequence of Homo sapiens clone  
Unpublished  
2 (bases 1 to 166458)  
Waterston, R.H.  
Direct Submission  
Submitted (10-NOV-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Nov 11, 2000 this sequence version replaced gi:11136834.

COMMENT

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
----- Project Information -----  
Center project name: H\_FH0516M14  
----- Summary Statistics -----  
Sequencing vector: M13; 100%  
Chemistry: Dye-primer ET; 100% of reads  
Chemistry: Dye-terminator Big Dye; 0% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 147705 bases at least Q40  
Consensus quality: 155070 bases at least Q30  
Consensus quality: 158378 bases at least Q20  
Insert size: 172000; agarose-fp  
Insert size: 164058; sum-of-contigs  
Quality coverage: 3.66 in Q20 bases; agarose-fp  
Quality coverage: 3.90 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 25 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1192: contig of 1192 bp in length  
\* 1193 1292: gap of unknown length  
\* 1293 2830: contig of 1538 bp in length  
\* 2831 2930: gap of unknown length  
\* 2931 4485: contig of 1555 bp in length  
\* 4486 4585: gap of unknown length  
\* 4586 7134: contig of 2549 bp in length  
\* 7135 7235: gap of unknown length  
\* 7235 9400: contig of 2166 bp in length  
\* 9401 9500: gap of unknown length  
\* 9501 11389: contig of 1889 bp in length  
\* 11390 11489: gap of unknown length  
\* 11490 13890: contig of 2401 bp in length  
\* 13891 13990: gap of unknown length  
\* 13991 16786: contig of 2796 bp in length  
\* 16787 16886: gap of unknown length  
\* 16887 20616: contig of 3730 bp in length  
\* 20617 20716: gap of unknown length  
\* 20717 23530: contig of 2814 bp in length  
\* 23531 23630: gap of unknown length  
\* 23631 26912: contig of 3282 bp in length  
\* 26913 27012: gap of unknown length  
\* 27013 29489: contig of 2477 bp in length  
\* 29490 29589: gap of unknown length  
\* 29590 32528: contig of 2939 bp in length  
\* 32529 32629: gap of unknown length  
\* 32630 38476: contig of 5847 bp in length  
\* 38476 38576: gap of unknown length  
\* 38576 43098: contig of 4523 bp in length  
\* 43099 43198: gap of unknown length  
\* 43199 46803: contig of 3605 bp in length  
\* 46804 46903: gap of unknown length

FEATURES  
Source

1. 166458  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="UL"  
/clone="RP13-516M14"  
1. 1192  
/note="assembly\_name:Contig10"  
1293. 2830  
/note="assembly\_name:Contig11"  
2931. 4485  
/note="assembly\_name:Contig12"  
4586. 7134  
/note="assembly\_name:Contig13"  
7235. 9400  
/note="assembly\_name:Contig14"  
9501. 11389  
/note="assembly\_name:Contig15"  
11490. 13890  
/note="assembly\_name:Contig16"  
13991. 16786  
/note="assembly\_name:Contig17"  
16887. 20616  
/note="assembly\_name:Contig18"  
20717. 23530  
/note="assembly\_name:Contig19"  
23631. 26912  
/note="assembly\_name:Contig20"  
27013. 29489  
/note="assembly\_name:Contig21"  
29590. 32528  
/note="assembly\_name:Contig22"  
clone\_end:SP6  
vector\_side:left  
32629. 38475  
/note="assembly\_name:Contig29"  
clone\_end:T7  
vector\_side:left  
38576. 43098  
/note="assembly\_name:Contig23"  
43199. 46803  
/note="assembly\_name:Contig24"  
46904. 51761  
/note="assembly\_name:Contig25"  
51862. 57363  
/note="assembly\_name:Contig26"  
57464. 65221  
/note="assembly\_name:Contig27"  
65322. 72371  
/note="assembly\_name:Contig28"  
72472. 80691  
/note="assembly\_name:Contig30"  
80792. 88618  
/note="assembly\_name:Contig31"  
88719. 109660  
/note="assembly\_name:Contig32"

```
misc_feature 109761..132756
/note="assembly_name:Contig33"
misc_feature 132857..166458
/note="assembly_name:Contig34"
BASE COUNT 37679 a 46105 c 45171 g 35080 t 2423 others
ORIGIN
Query Match 87.6%; Score 18.4; DB 2; Length 166458;
Best Local Similarity 95.0%; Pred. No. 90;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 gcatggagatgttgcagc 21
||||| |||||||||
Db 156111 GCATGCTAGATGTCGCAGC 156092

RESULT 14
AC009564/c
DEFINITION Homo sapiens chromosome 15 clone RP11-413M23 map 15, WORKING DRAFT
SEQUENCE, 36 unordered pieces.
AC009564
VERSION AC009564.5 GI:8072465
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 184090)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-413M23
Unpublished
2 (bases 1 to 184090)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Collangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Testaye,S., Torruella-Miller,J., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (27-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:7637235.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1692
Center clone name: 413_M23
----- Summary Statistics
Sequencing vector: M13; M77815; 96% of reads
Sequencing vector: Plasmid; n/a; 0.0% of reads
4.25090720580612Chemistry: Dye-primer-amersham; 4% of reads
Chemistry: Dye-terminator Big Dye; 96% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 151999 bases at least Q40
Consensus quality: 164409 bases at least Q30
Consensus quality: 171030 bases at least Q20

Insert size: 182000; agarose-fp
Insert size: 180590; sum-of-contigs
Quality covera.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1060: contig of 1060 bp in length
* 1061 1160: gap of 100 bp
* 1161 2375: contig of 1215 bp in length
* 2376 2475: gap of 100 bp
* 2476 3802: contig of 1327 bp in length
* 3803 3902: gap of 100 bp
* 3903 5236: contig of 1334 bp in length
* 5237 5336: gap of 100 bp
* 5337 7037: contig of 1701 bp in length
* 7038 7137: gap of 100 bp
* 7138 8383: contig of 1246 bp in length
* 8384 8483: gap of 100 bp
* 8484 9908: contig of 1425 bp in length
* 9909 10008: gap of 100 bp
* 10009 11137: contig of 1129 bp in length
* 11138 11237: gap of 100 bp
* 11238 13218: contig of 1981 bp in length
* 13219 13318: gap of 100 bp
* 13319 16067: contig of 2749 bp in length
* 16068 16167: gap of 100 bp
* 16168 18833: contig of 2666 bp in length
* 18834 18933: gap of 100 bp
* 18934 22378: contig of 3445 bp in length
* 22379 22478: gap of 100 bp
* 22479 24946: contig of 2468 bp in length
* 24947 25046: gap of 100 bp
* 25047 26306: contig of 1260 bp in length
* 26307 26406: gap of 100 bp
* 26407 29679: contig of 3273 bp in length
* 29680 29779: gap of 100 bp
* 29780 33599: contig of 3820 bp in length
* 33600 36599: gap of 100 bp
* 33700 36502: contig of 2803 bp in length
* 36503 36602: gap of 100 bp
* 36603 39914: contig of 3312 bp in length
* 39915 40014: gap of 100 bp
* 40015 42987: contig of 2974 bp in length
* 42988 43087: gap of 100 bp
* 43088 46873: contig of 3786 bp in length
* 46874 46973: gap of 100 bp
* 46974 50544: contig of 3571 bp in length
* 50545 50644: gap of 100 bp
* 50645 53900: contig of 3256 bp in length
* 53901 54000: gap of 100 bp
* 54001 58793: contig of 4793 bp in length
* 58794 58993: gap of 100 bp
* 58994 64521: contig of 5628 bp in length
* 64522 64621: gap of 100 bp
* 64622 69423: contig of 4802 bp in length
* 69424 69523: gap of 100 bp
* 69524 75858: contig of 6335 bp in length
* 75859 75958: gap of 100 bp
* 75959 82538: contig of 6580 bp in length
* 82539 82638: gap of 100 bp
* 82639 89982: contig of 7344 bp in length
* 89983 90082: gap of 100 bp
* 90083 96628: contig of 6546 bp in length
* 96629 96728: gap of 100 bp
* 96729 106892: contig of 10164 bp in length
* 106893 106992: gap of 100 bp
* 106993 118329: contig of 11337 bp in length
* 118330 118429: gap of 100 bp
```

\* 118430 130262: contig of 11833 bp in length  
 \* 130263 130362: gap of 100 bp  
 \* 130363 141462: contig of 11100 bp in length  
 \* 141463 141562: gap of 100 bp  
 \* 141563 154056: contig of 12494 bp in length  
 \* 154057 154156: gap of 100 bp  
 \* 154157 164998: contig of 10842 bp in length  
 \* 164999 165098: gap of 100 bp  
 \* 165099 184090: contig of 18992 bp in length.

## FEATURES

source  
 1. 184090  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="15"  
 /map="15"  
 /clone="RP11-413M23"  
 /clone.lib="RPC1-11 Human Male BAC"  
 1. 1060  
 /note="assembly\_fragment"  
 1161. 2375  
 /note="assembly\_fragment"  
 2476. 3802  
 /note="assembly\_fragment"  
 3903. 5236  
 /note="assembly\_fragment"  
 5337. 7037  
 /note="assembly\_fragment"  
 7138. 8383  
 /note="assembly\_fragment"  
 8484. 9908  
 /note="assembly\_fragment"  
 10009. 11137  
 /note="assembly\_fragment"  
 11238. 13218  
 /note="assembly\_fragment"  
 13319. 16067  
 /note="assembly\_fragment"  
 16168. 18833  
 /note="assembly\_fragment"  
 18934. 22378  
 /note="assembly\_fragment"  
 22479. 24946  
 /note="assembly\_fragment"  
 25047. 26306  
 /note="assembly\_fragment"  
 clone\_end:T7  
 vector\_side:left"  
 26407. 29679  
 /note="assembly\_fragment"  
 29780. 33599  
 /note="assembly\_fragment"  
 33700. 36502  
 /note="assembly\_fragment"  
 36803. 39914  
 /note="assembly\_fragment"  
 40015. 42987  
 /note="assembly\_fragment"  
 clone\_end:SP6  
 vector\_side:right"  
 43088. 46873  
 /note="assembly\_fragment"  
 46974. 50544  
 /note="assembly\_fragment"  
 50645. 53900  
 /note="assembly\_fragment"  
 54001. 58793  
 /note="assembly\_fragment"  
 58894. 64521  
 /note="assembly\_fragment"  
 64622. 69423  
 /note="assembly\_fragment"  
 69524. 75858  
 /note="assembly\_fragment"

Query Match 87.6%; Score 18.4; DB 2: Length 184090;  
 Best Local Similarity 95.0%; Pred. No. 89;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 agcattggagatgttgacg 20  
 ||| ||||| ||||| |||||  
 Db 83816 AGCTGGGAGATGTTGGCAG 83797

## RESULT 15

AC099069/c  
 LOCUS  
 DEFINITION  
 AC099069  
 VERSION  
 AC099069.3 GI:17943533  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Rattus norvegicus  
 Norway rat.  
 AC099069  
 Rattus norvegicus chromosome Rf1 clone CH230-1006, WORKING DRAFT  
 SEQUENCE, 27 unordered pieces.

## REFERENCE

1 (bases 1 to 217028)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J.,  
 Benton,J., Bimaga,K., Blank,burg,K., Bonnin,D., Bouck,J.,  
 Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
 Davy-Carroll,L., Dederich,D.A., Delaney,K.K., Delgado,O.,  
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
 Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,  
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Haves,A.,  
 Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Hollaway,C.,  
 Hollins,B., Homsif., Howard,S., Huber,J., Huiy,K.S., Hume,J.,  
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,  
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,  
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
 Loulsegedi,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,  
 Martinez,E., Massey,E., Massey,E., McLeod,M.P., Meador,M.,  
 Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S.,  
 Ogah,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,  
 Peery,J., Perez,L., Peters,L., Pichens,R., Primus,E., Pu,L.,  
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,  
 Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshkari,N.,  
 Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,  
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,  
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,  
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,  
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,  
 Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S.,  
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstein,G. and Gibbs,R.

## TITLE

JOURNAL  
 Direct Submission

REFERENCE  
 2 (bases 1 to 217028)

Worley,K.C.  
 Direct Submission

## AUTHORS

TITLE  
 JOURNAL  
 Submitted (09-NOV-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Dec 20, 2001 this sequence version replaced gi:17062212.

## COMMENT

----- Genome Center

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GDU1  
Center clone name: CH230-1006  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329First call to  
findPhrapList

Consensus quality: 198668 bases at least Q40  
Consensus quality: 202349 bases at least Q30  
Consensus quality: 205127 bases at least Q20  
Estimated insert size: 206321; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-fp estimation  
Quality coverage: 4.2x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 27 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
26363: contig of 26363 bp in length  
26364: gap of unknown length  
26464: contig of 19829 bp in length  
46292: contig of unknown length  
46293: gap of unknown length  
46393: contig of 20296 bp in length  
66688: gap of unknown length  
66788: gap of unknown length  
82769: contig of 15981 bp in length  
82770: gap of unknown length  
82870: contig of 15552 bp in length  
98421: contig of 13616 bp in length  
98522: gap of unknown length  
112138: contig of 9377 bp in length  
112238: gap of unknown length  
121615: gap of unknown length  
121714: contig of 10173 bp in length  
131887: gap of unknown length  
131888: contig of 12780 bp in length  
144767: gap of unknown length  
144768: gap of unknown length  
144867: gap of unknown length  
152164: contig of 7297 bp in length  
152264: gap of unknown length  
15265: gap of unknown length  
15265: contig of 6333 bp in length  
158597: gap of unknown length  
158598: gap of unknown length  
158697: gap of unknown length  
164577: contig of 5880 bp in length  
164578: gap of unknown length  
164678: gap of unknown length  
170987: contig of 6309 bp in length  
170987: gap of unknown length  
171087: contig of 5373 bp in length  
176459: gap of unknown length  
176460: gap of unknown length  
176560: contig of 5716 bp in length  
182275: gap of unknown length  
182375: gap of unknown length  
182376: contig of 6964 bp in length  
189339: gap of unknown length  
189340: gap of unknown length  
189440: contig of 4495 bp in length  
193935: gap of unknown length  
194035: contig of 5662 bp in length  
199697: gap of unknown length  
199796: gap of unknown length  
202752: contig of 2956 bp in length  
202753: gap of unknown length  
202852: gap of unknown length  
202853: contig of 3136 bp in length  
205988: gap of unknown length  
205989: gap of unknown length  
206088: gap of unknown length  
207601: contig of 1513 bp in length  
207602: gap of unknown length  
207602: gap of unknown length  
209268: contig of 1567 bp in length  
209269: gap of unknown length  
209369: contig of 1415 bp in length  
210783: contig of 1415 bp in length  
210784: gap of unknown length

\* 210884 212298: contig of 1415 bp in length  
\* 212399 212399: gap of unknown length  
\* 214566 214566: contig of 2168 bp in length  
\* 214567 214567: gap of unknown length  
\* 215779 215779: contig of 1113 bp in length  
\* 215780 215780: gap of unknown length  
\* 215880 217028: contig of 1149 bp in length.  
FEATURES  
Location/Qualifiers  
source  
1. 217028  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/chromosome="Rfl"  
/clone="CH230-1006"  
BASE COUNT 65529 a 43513 c 43420 g 61834 t 2732 others  
ORIGIN  
Query Match. 87.6%; Score 18.4; DB 2: Length 217028;  
Best Local Similarity 95.0%; Pred. No. 89;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 agcatggagagatgtgcag 20  
|||||||  
Db 181189 ACCATGGGAGATGATGGCAG 181170

Search completed: May 18, 2002, 14:59:29  
Job time: 8266 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:22:25 ; Search time 8624.33 Seconds  
(without alignments)

46.950 Million cell updates/sec

Title: US-09-623-329-26

Perfect score: 30

Sequence: 1 ctggttaactgggcccaggagagagca 30

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: gb\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
C 1	21.6	72.0	313	9	BB118220	BB118220 BB118220
2	21.6	72.0	1061	12	CNS02C0H	AL190538 Tetraodon
3	20.6	68.7	771	12	AZ175624	AZ175624 SP.0133-A
4	20.6	68.7	985	10	BG854980	BG854980 1024041D0
5	20.4	68.0	368	9	AA291011	AA291011 zs45gll.r
6	20.4	68.0	409	10	BE554311	BE554311 ur40g10.y
7	20.4	68.0	647	12	AQ576204	AQ576204 nbxb0088D
8	20.4	68.0	1105	10	BF158832	BF158832 601766183
C 9	20	66.7	252	9	AV282744	AV282744 AV282744
C 10	20	66.7	303	4	B8708681	B8708681 Mus muscu
C 11	20	66.7	321	9	B8204557	B8204557 BB204557
12	20	66.7	563	12	BH332455	BH332455 CH230-189
13	20	66.7	688	12	AG059331	AG059331 Pan trogl
14	20	66.7	697	12	AG109518	AG109518 Pan trogl
C 15	20	66.7	785	10	BG221191	BG221191 RST40994
C 16	20	66.7	901	10	BG743995	BG743995 602722767
C 17	19.8	66.0	324	10	BF389515	BF389515 UI-R-CJ0-

18	19.8	66.0	446	10	BG813204	BG813204 da172c06.
19	19.8	66.0	526	5	B1314422	B1314422 da101b12.
20	19.8	66.0	596	10	BJ058360	BJ058360 BJ058360
21	19.8	66.0	599	10	BJ064020	BJ064020 BJ064020
22	19.8	66.0	608	10	BJ057732	BJ057732 BJ057732
23	19.6	65.3	182	12	BH289971	BH289971 CH230-172
24	19.6	65.3	531	12	AQ080462	AQ080462 HS_3149_B
25	19.6	65.3	852	9	AV753136	AV753136 AV753136
26	19.6	65.3	896	12	CNS03C1E	AL237227 Tetraodon
C 27	19.4	64.7	318	10	BF013157	BF013157 ux97g10.y
C 28	19.4	64.7	377	9	AI320586	AI320586 C7c02mm.f
C 29	19.4	64.7	390	9	AI320678	AI320678 c8g03mm.f
C 30	19.4	64.7	393	9	AW725014	AW725014 f9d06nm.f
C 31	19.4	64.7	469	9	AW716081	AW716081 g9a05nm.f
C 32	19.4	64.7	478	9	AW714914	AW714914 gla01nm.f
C 33	19.4	64.7	487	12	AQ657417	AQ657417 Sheared D
C 34	19.4	64.7	493	10	T99251	T99251 ye63h04.sl
C 35	19.4	64.7	506	10	B1012416	B1012416 QV0-EN005
C 36	19.4	64.7	543	9	AI320683	AI320683 c8g11nm.f
C 37	19.4	64.7	546	12	TAG62C1IP	AL450040 T. brucei
C 38	19.4	64.7	559	9	AI316861	AI316861 ul35b05.y
C 39	19.4	64.7	625	9	AA149466	AA149466 zo29b05.s
C 40	19.4	64.7	633	10	BE724660	BE724660 894077H03
C 41	19.4	64.7	676	12	AZ319231	AZ319231 JM0038C13
C 42	19.4	64.7	711	12	AG111618	AG111618 Pan trogl
C 43	19.4	64.7	743	9	AW964214	AW964214 EST376287
C 44	19.4	64.7	879	10	BF161132	BF161132 601768072
C 45	19.2	64.0	288	9	BB479448	BB479448 BB479448

#### ALIGNMENTS

RESULT 1

BB118220/c

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BB118220 313 bp mRNA linear EST 27-JUN-2000  
BB118220 RIKEN full-length enriched, adult male urinary bladder Mus  
musculus CDNA clone 9530065A02 3', mRNA sequence.

BB118220  
BB118220  
BB118220.1 GI:8770788

EST.  
Mus musculus  
house mouse.

Eukaryota: Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 313)

Kunno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci,  
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,  
Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,  
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,  
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,  
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,  
Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata,  
Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,  
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya,  
T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I.,  
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino,  
M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)

Unpublished (2000)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki

N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Thermotabilization and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length







Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 409)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index  
 Unpublished (1997)  
 Other\_Ests: ur40g10.x1

Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs@mail.nih.gov  
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 MGI:105558

Seq primer: -40RP from Gibco  
 High quality sequence stop: 353.

#### FEATURES

Location/Qualifiers  
 1. .409  
 /organism="Mus musculus"  
 /strain="FVB/N-3"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3152802"  
 /tissue\_type="tumor, biopsy sample"  
 /dev\_stage="5 months"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"  
 70 a 138 c 129 g 72 t

Query Match 68.0%; Score 20.4; DB 10; Length 409;  
 Best Local Similarity 80.0%; Pred. No. 1.9e+02;  
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ctggtttaaactggccaggagagca 30  
 ||||| | ||||| ||||| |||||

Db 261 CTGTCAGGCTGGCCAGGCAAGGACCA 290

RESULT 7  
 AQ576204  
 LOCUS  
 DEFINITION  
 nbxb0088D02f CUGI Rice BAC Library Oryza sativa genomic clone

ACCESSION AQ576204

VERSION AQ576204.1

KEYWORDS GSS.

SOURCE Oryza sativa.

ORGANISM Oryza sativa.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 647)

Wing, R.A. and Dean, R.A.

A BAC End Sequencing Framework to Sequence the Rice Genome

Unpublished (1998)

Contact: Wing RA

Clemson University

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: [rwing@clemson.edu](mailto:rwing@clemson.edu)

Seq primer: TAATAGGACTCATATAGG

Class: BAC ends

High quality sequence start: 8

High quality sequence stop: 124.

Location/Qualifiers

1. .647

/organism="Oryza sativa"

#### FEATURES

source

/strain="Japonica"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:4530"  
 /clone="nbxb0088D02f"  
 /clone\_lib="CUGI Rice BAC Library"  
 /tissue\_type="Leaf"  
 /lab\_host="E. coli DH10B"  
 /note="Vector: pBelOBAC11; Site\_1: HindIII; Site\_2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation of a particular sequence with a probability of 99.9%. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."  
 226 a 115 c 155 g 151 t

#### BASE COUNT

ORIGIN

Query Match 68.0%; Score 20.4; DB 12; Length 647;  
 Best Local Similarity 80.0%; Pred. No. 2.2e+02;  
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ctggtttaaactggccaggagagca 30

||||| | ||||| ||||| |||||

Db 518 CTGTTTACACTGCCCCGGAGAGGAA 547

#### RESULT

8

BF158832

LOCUS

DEFINITION

BF158832

VERSION

KEYWORDS

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 1105)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAW9180 row: e column: 22

High quality sequence stop: 713.

Location/Qualifiers

1. .1105

/organism="Mus musculus"

/strain="C2EHC II (feral)"

/db\_xref="taxon:10090"

/clone="IMAGE:3982077"

/clone\_lib="NCI\_CGAP\_Lu29"

#### FEATURES

source





cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATCTCGAGTTAAATTAAATCCCGCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 94 a 70 c 44 g 113 t

ORIGIN

Query Match 66.7%; Score 20; DB 9; Length 321;  
Best Local Similarity 82.1%; Pred. No. 2.7e+02;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 ggtttaaactggccaggagagagca 30

DB 64 GTTTAAAGGGGGCCAGGAGATTGCA 37

RESULT 12

LOCUS BH332455 563 bp DNA linear GSS 03-DEC-2001

DEFINITION CH230-189M7.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone

ACCESSION BH332455

VERSION BH332455.1 GI:17263169

KEYWORDS GSS.

SOURCE Norway rat.

ORGANISM

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 563)

AUTHORS Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shwartsbeyn, A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P., and Fraser, C.M.

Rat BAC End sequences from Library CHORI-230 ECORI segment

Unpublished (1999)

Other\_GSSs: CH230-189M7.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

7112 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230

(<http://www.chori.org/bacpac/rat230.htm>). For BAC library

availability, please contact Pieter de Jong ([pdejong@mail.cho.org](mailto:pdejong@mail.cho.org)).

Clones may be purchased from BACPAC Resources

([http://www.chori.org/bacpac/orering\\_information.htm](http://www.chori.org/bacpac/orering_information.htm)). BAC end

page: [http://www.tigr.org/tdb/bac\\_ends/rat/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html)

Plate: 189 row: M column: 7

Seq primer: T7

Class: BAC ends.

FEATURES

Source

Location/Qualifiers

1..563

/organism="Rattus norvegicus"

/strain="BN/SSNHsd/MCW"

/db\_xref="taxon:10116"

/clone="CH230-189M7"

/clone\_lib="CHORI-230 Segment 1"

/sex="Female"

/cell\_type="Brain"

/note="Vector: pTARBAC2.1; Site\_1: EcoRI; Site\_2: EcoRI;

CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by

Pieter de Jong"

143 a 113 c 109 g 198 t

BASE COUNT

ORIGIN

Query Match 66.7%; Score 20; DB 12; Length 563;

Best Local Similarity 82.1%; Pred. No. 3.1e+02;

Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ctggtttaaactggccaggagagag 28

DB 290 CTGTAGAACTGGTCCAGGAAGAG 317

RESULT 13

LOCUS AG059331

DEFINITION Pan troglodytes DNA, clone: PTB-046H21.R, genomic survey sequence.

ACCESSION AG059331

VERSION AG059331.1 GI:16610561

KEYWORDS GSS; GSS (genome survey sequence).

SOURCE Pan troglodytes male lymphoblast DNA, clone\_lib:PTB Chimpanzee Male

BAC Library clone:PTB-046H21.K.

ORGANISM

Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE

AUTHORS

1 (sites)

Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,

Totoki, Y., Watanabe, H. and Sakaki, Y.

BAC end sequences of Library PTB

Unpublished

2 (bases 1 to 688)

Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,

Totoki, Y., Watanabe, H. and Sakaki, Y.

Direct Submission

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail: [chimbse@sc.riken.go.jp](mailto:chimbse@sc.riken.go.jp), URL: <http://ngp.gsc.riken.go.jp/>,

Tel: 81-45-503-9111, Fax: 81-45-503-9170)

Clones are derived from the chimpanzee BAC library PTB This BAC end

was generated during the R&D process and may have higher chance of

clone tracking errors.

PRIMERS

Sequencing: M13Rev

LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1..688

/organism="Pan troglodytes"

/db\_xref="taxon:9598"

/clone="PTB-046H21.R"

/sex="male"

/cell\_type="lymphoblast"

/clone\_lib="PTB Chimpanzee Male BAC Library"

193 a 146 c 190 g 158 t

others

BASE COUNT

ORIGIN

Query Match 66.7%; Score 20; DB 12; Length 688;

Best Local Similarity 82.1%; Pred. No. 3.2e+02;

Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ctggtttaaactggccaggagagag 28

DB 314 CTGGCTTAGAGTGGCCCGAGAGAGAG 341

RESULT 14

LOCUS AG109518

DEFINITION Pan troglodytes DNA, clone: PTB-114P11.R, genomic survey sequence.

ACCESSION AG109518

VERSION AG109518.1 GI:16730037

KEYWORDS GSS; GSS (genome survey sequence).

SOURCE Pan troglodytes male lymphoblast DNA, clone\_lib:PTB Chimpanzee Male

BAC Library clone:PTB-114P11.R.

ORGANISM Pan troglodytes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 REFERENCE 1 (sites)  
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,  
 Totoki, Y., Watanabe, H. and Sakaki, Y.  
 TITLE BAC end sequences of Library PTB  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 697)  
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,  
 Totoki, Y., Watanabe, H. and Sakaki, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 1-7-22 Suenro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail: chimbesgc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,  
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
 COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end  
 was generated during the R&D process and may have higher chance of  
 clone tracking errors.  
 PRIMERS  
 Sequencing: M13Rev  
 LIBRARY  
 Vector : pKS145  
 R.Site 1 : SacI  
 R.Site 2 : SacI  
 Location/Qualifiers  
 1..697  
 /organism="Pan troglodytes"  
 /db\_xref="taxon:9598"  
 /clone="PTB-114P11.R"  
 /sex="male"  
 /cell\_type="lymphoblast"  
 /clone\_lib="PTB Chimpanzee Male BAC Library"  
 BASE COUNT 163 a 189 c 135 g 206 t 4 others  
 ORIGIN

Query Match 66.7%; Score 20; DB 12; Length 697;  
 Best Local Similarity 82.1%; Pred. NO. 3.2e+02;  
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 3 gggttaaacctggccagagagagagca 30  
 ||||| ||||| ||||| ||||| |||||  
 Db 103 GGATTACTGTGCCAGCTGAGAGCA 130

RESULT 15  
 BG221191/c  
 LOCUS  
 DEFINITION RST40994 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BG221191  
 VERSION BG221191.1 GI:13747212  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 785)  
 AUTHORS Harrington, J. J., Sherf, B., Rundlett, S., Jackson, P. D., Perry, R.,  
 Cain, S., Leventhal, C., Thornton, M., Ramachandran, K., Whittington, J.,  
 Lerner, L., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith,  
 E., Veloso, N., Klika, A., Hess, J., Cothren, K., Lo, K., Offenbacher,  
 J., Danzig, J. and Ducar, M.  
 TITLE Creation of genome-wide protein expression libraries using random  
 activation of gene expression  
 JOURNAL Nat. Biotechnol. 19 (5), 440-445 (2001)  
 MEDLINE 2127151  
 COMMENT Contact: Scott J. Cain  
 Athersys, Inc.  
 3201 Carnegie Ave, Cleveland, OH 44115, USA  
 Tel: 216 431 9900  
 Fax: 216 361 9596

Email: scaine@atersys.com  
 High quality sequence stop: 492.  
 Location/Qualifiers  
 1..785  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Athersys RAGE Library"  
 /cell\_line="HT1080"  
 /note="See 'Creation of Genome-wide Protein Expression'  
 Libraries using Random Activation of Gene Expression',  
 Nature Biotechnology, in press. Note that even though the  
 cell type indicated is HT1080, since a random activation  
 method was used, these sequence tags are not necessarily  
 expressed in HT1080 under normal circumstances."  
 BASE COUNT 176 a 247 c 164 g 195 t 3 others  
 ORIGIN

Query Match 66.7%; Score 20; DB 10; Length 785;  
 Best Local Similarity 82.1%; Pred. NO. 3.3e+02;  
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 2 tggtttaaacctggccagagagagagc 29  
 ||||| ||||| ||||| ||||| |||||  
 Db 78 TGGTTTCAAAATGGGCTCAGGAGTGAAGC 51

Search completed: May 18, 2002, 17:22:28  
 Job time: 14699 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:26:43 ; Search time 250.21 Seconds  
(without alignments)  
29.451 Million cell updates/sec

Title: US-09-623-329-26  
Perfect score: 30  
Sequence: 1 ctggttaactggccaggagagagca 30

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	17.8	59.3	35524	3	US-08-923-137-1
2	17.4	58.0	3153	4	US-09-175-928-9
3	17.2	57.3	672	2	US-08-446-875-13
C 4	17.2	57.3	784	1	US-08-392-731-1
C 5	17.2	57.3	1186	1	US-08-368-236-2
C 6	17.2	57.3	2874	4	US-09-179-558-54
C 7	17.2	57.3	3059	4	US-09-179-558-61
C 8	17.2	57.3	3202	4	US-09-287-354-1
C 9	17.2	57.3	3363	3	US-09-258-349-1
C 10	17.2	57.3	3791	5	PCT-US96-01735-5
C 11	17.2	57.3	28958	1	US-08-258-261B-6
C 12	17.2	57.3	28958	1	US-08-456-837-6
C 13	17.2	57.3	28958	1	US-08-457-342-6
C 14	17.2	57.3	28958	1	US-08-457-645A-6
C 15	17.2	57.3	28958	1	US-08-458-076A-6
C 16	17.2	57.3	28958	1	US-08-764-233A-4
C 17	17.2	57.3	28958	1	US-08-457-335A-6
C 18	17.2	57.3	28958	1	US-08-729-214-6
C 19	17.2	57.3	28958	3	US-09-028-934-6
C 20	17.2	57.3	49377	1	US-08-764-233A-1
C 21	17	56.7	754	4	US-09-328-111-447
C 22	17	56.7	954	2	US-08-555-722-1
C 23	17	56.7	954	4	US-09-384-301-1
C 24	16.8	56.0	354	3	US-08-513-968-79
C 25	16.8	56.0	1941	2	US-09-008-960-2
C 26	16.8	56.0	1941	3	US-09-368-240-2
C 27	16.8	56.0	1941	4	US-09-468-702-2

Sequence 1, Appli  
Sequence 14, Appli  
Sequence 14, Appli  
Sequence 12, Appli  
Sequence 12, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 31, Appli  
Sequence 32, Appli  
Sequence 14, Appli  
Sequence 16, Appli  
Sequence 18, Appli  
Sequence 12, Appli  
Sequence 14, Appli  
Sequence 16, Appli  
Sequence 18, Appli  
Sequence 19, Appli  
Sequence 1, Appli  
Sequence 14, Appli  
Sequence 14, Appli  
Sequence 12, Appli  
Sequence 12, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 31, Appli  
Sequence 32, Appli  
Sequence 14, Appli  
Sequence 16, Appli  
Sequence 18, Appli  
Sequence 12, Appli  
Sequence 14, Appli  
Sequence 16, Appli  
Sequence 18, Appli  
Sequence 19, Appli

## ALIGNMENTS

RESULT 1  
US-08-923-137-1/C  
; Sequence 1, Application US/08923137  
; Patent No. 6083716  
; GENERAL INFORMATION:  
; APPLICANT: Wilson, James M.  
; APPLICANT: Farina, Steven F.  
; APPLICANT: Fisher, Krishna J.  
; TITLE OF INVENTION: Chimpanzee Adenovirus Vectors  
; NUMBER OF INVENTIONS: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howson and Howson  
; STREET: Spring House Corporate Cntr., P.O. Box 457  
; CITY: Spring House  
; STATE: Pennsylvania  
; COUNTRY: United States of America  
; ZIP: 19477  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/923,137  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/024,700  
; FILING DATE: 06-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bak, Mary E.  
; REGISTRATION NUMBER: 31,215  
; REFERENCE/DOCKET NUMBER: GNPVN.021CIP1USA  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-540-9200  
; TELEFAX: 215-540-5818  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 35524 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: CDNA  
US-08-923-137-1

Query Match 59.3%; Score 17.8; DB 3; Length 35524;  
Best Local Similarity 75.9%; Pred. No. 55;  
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;





HYPOTHETICAL: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Sorangium cellulosum  
 IMMEDIATE SOURCE:  
 CLONE: Cosmid clone p98/1  
 FEATURE:  
 NAME/KEY: misc.feature  
 LOCATION: 1..784  
 OTHER INFORMATION: /product= "Constituent of the  
 OTHER INFORMATION: 'soraphen gene cluster'."  
 US-08-392-731-1

Query Match 57.3%; Score 17.2; DB 1; Length 784;  
 Best Local Similarity 86.4%; Pred. No. 54;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 9 aactggggccaggagagagca 30  
 ||| ||||| ||||| |||||  
 DB 371 AGCTCCGCCAGGAGGAGCA 350

RESULT 5  
 US-08-368-236-2/C  
 Sequence 2, Application US/08368236  
 Patent No. 5804408  
 GENERAL INFORMATION:  
 APPLICANT: Hagiwara, et al.  
 TITLE OF INVENTION: A METHOD FOR EXPRESSING POLYPEPTIDES  
 NUMBER OF SEQUENCES: 2  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sherman and Shalloway  
 STREET: 413 N. Washington Street  
 CITY: Alexandria  
 STATE: VA  
 COUNTRY: USA  
 ZIP: 22314

COMPUTER READABLE FORM:  
 MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.4 MB STORAGE  
 COMPUTER: IBM  
 OPERATING SYSTEM: DOS 5.1  
 SOFTWARE: Wordperfect 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/368,236  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/941,139  
 FILING DATE: No. 5804408ember 13, 1992  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Richard Steinberg.  
 REGISTRATION NUMBER: 26,588  
 REFERENCE/DOCKET NUMBER: S-2336  
 TELEPHONE: (703) 549-2282  
 TELEFAX: (703) 836-0106  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1186  
 TYPE: Nucleic Acid  
 STRANDEDNESS: Single  
 TOPOLOGY: Linear  
 ANTI-SENSE: NO  
 US-08-368-236-2

Query Match 57.3%; Score 17.2; DB 1; Length 1186;  
 Best Local Similarity 73.3%; Pred. No. 58;  
 Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 ctggtttaaactggggccaggagagca 30

DB 618 CGGGTTGAATGAGGCCCTGCAGAGTGCA 589  
 ||||| ||| ||||| ||||| ||||| |||||  
 RESULT 6  
 US-09-179-558-54/c  
 Sequence 54, Application US/09179558  
 Patent No. 6180612  
 GENERAL INFORMATION:  
 APPLICANT: Hockensmith, Joel W.  
 APPLICANT: Muthuswami, Rohini  
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
 TARGETING DNA METABOLIC PROCESSES USING  
 TITLE OF INVENTION: AMINOGLYCOSIDE DERIVATIVES  
 NUMBER OF SEQUENCES: 66  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: PENNIE & EDMONDS LLP  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10036-2711  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/179,558  
 FILING DATE: 27-OCT-1998  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: U.S. 09/060,470  
 FILING DATE: 15-APR-1998  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: U.S. 60/063,898  
 FILING DATE: 31-OCT-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Coruzzi, Laura A  
 REGISTRATION NUMBER: 30,742  
 REFERENCE/DOCKET NUMBER: 9426-005-999  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212)7909090  
 TELEFAX: (212)8699741  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 54:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2874 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: Other  
 US-09-179-558-54

Query Match 57.3%; Score 17.2; DB 4; Length 2874;  
 Best Local Similarity 73.3%; Pred. No. 67;  
 Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ctggtttaaactggggccaggagagca 30  
 ||| | || ||||| ||||| |||||  
 DB 1033 CTCGCTCGAAGCGGGCCGGGAGATGAGCA 1004

RESULT 7  
 US-09-179-558-61/c  
 Sequence 61, Application US/09179558  
 Patent No. 6180612  
 GENERAL INFORMATION:  
 APPLICANT: Hockensmith, Joel W.  
 APPLICANT: Muthuswami, Rohini  
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
 TARGETING DNA METABOLIC PROCESSES USING

;; TITLE OF INVENTION: AMINOGLYCOSIDE DERIVATIVES  
;; NUMBER OF SEQUENCES: 66  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: PENNIE & EDMONDS LLP  
;; STREET: 1155 Avenue of the Americas  
;; CITY: New York  
;; STATE: NY  
;; COUNTRY: USA  
;; ZIP: 10036-2711  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FastSeq Version 2.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/179,558  
;; FILING DATE: 27-OCT-1998  
;; CLASSIFICATION: 514  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: U.S. 09/060,470  
;; FILING DATE: 15-APR-1998  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: U.S. 60/063,898  
;; FILING DATE: 31-OCT-1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Coruzzi, Laura A  
;; REGISTRATION NUMBER: 30,742  
;; REFERENCE/DOCKET NUMBER: 9426-005-999  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212)7909090  
;; TELEFAX: (212)8699741  
;; TELEX: 66141 PENNIE  
;; INFORMATION FOR SEQ ID NO: 61:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 3059 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: Other  
US-09-179-558-61

Query Match 57.3%; Score 17.2; DB 4; Length 3059;  
Best Local Similarity 73.3%; Pred. No. 68;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ctggtttaaactggccaggagagagca 30  
||| | || ||||| ||||| |||||  
Db 1218 CTGCTGAAGC\*GGCCGGAGATGAGCA 1189

RESULT 8  
US-09-287-354-1/c  
;; Sequence 1, Application US/09287354  
;; Patent No. 6348348  
;; GENERAL INFORMATION:  
;; APPLICANT: THOMPSON, Catherine C.  
;; TITLE OF INVENTION: HUMAN HAIRLESS GENE AND PROTEIN  
;; FILE REFERENCE: Thompson-20263/0243435  
;; CURRENT APPLICATION NUMBER: US/09/287,354  
;; CURRENT FILING DATE: 1999-04-07  
;; EARLIER APPLICATION NUMBER: US 60/080,888  
;; EARLIER FILING DATE: 1998-04-07  
;; NUMBER OF SEQ ID NOS: 10  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 1  
;; LENGTH: 3202  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-287-354-1

Query Match 57.3%; Score 17.2; DB 4; Length 3202;

Best Local Similarity 73.3%; Pred. No. 68;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
Qy 1 ctggtttaaactggccaggagagagca 30  
||| | || ||||| ||||| |||||  
Db 354 CTGGGTGACAGCGGCTCAGGAGGGGCA 325

RESULT 9  
US-09-258-349-1  
;; Sequence 1, Application US/09258349  
;; Patent No. 6146878  
;; GENERAL INFORMATION:  
;; APPLICANT: Kumagai, Hidehiko  
;; APPLICANT: Suzuki, Hideyuki  
;; TITLE OF INVENTION: TYROSINE REPRESSOR GENE OF A BACTERIUM BELONGING TO THE  
;; TITLE OF INVENTION: GENUS ERWINIA  
;; FILE REFERENCE: OP819-US  
;; CURRENT APPLICATION NUMBER: US/09/258,349  
;; CURRENT FILING DATE: 1999-02-26  
;; EARLIER APPLICATION NUMBER: JP 10-48070  
;; EARLIER FILING DATE: 1998-02-27  
;; NUMBER OF SEQ ID NOS: 8  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 1  
;; LENGTH: 3363  
;; TYPE: DNA  
;; ORGANISM: Erwinia herbicola  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (442)..(2004)  
US-09-258-349-1

Query Match 57.3%; Score 17.2; DB 3; Length 3363;  
Best Local Similarity 73.3%; Pred. No. 69;  
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ctggtttaaactggccaggagagagca 30  
||| | || ||||| ||||| |||||  
Db 2339 ctggtttaaacttgcgcaccgaagagagca 2368

RESULT 10  
PCT-US96-01735-5/c  
;; Sequence 5, Application PC/TUS9601735  
;; GENERAL INFORMATION:  
;; APPLICANT: Marks, Andrew R.  
;; TITLE OF INVENTION: HUMAN T CELL INOSITOL 1,4,5,-TRISPHOSPHATE  
;; TITLE OF INVENTION: RECEPTOR  
;; NUMBER OF SEQUENCES: 8  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond  
;; STREET: 30 Rockefeller Plaza  
;; CITY: New York  
;; STATE: NY  
;; COUNTRY: USA  
;; ZIP: 10112-0228  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FastSeq Version 1.5  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US96/01735  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/386,039  
;; FILING DATE: 09-FEB-1995  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Kole, Lisa B

TELEFAX: 919-541-8689

; MOLECULE TYPE: DNA (genomic)

;  
;  
;  
US-08-456-837-6

Query Match 57.3%; Score 17.2; DB 1; Length 28958;  
Best Local Similarity 86.4%; Pred. No. 97;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 9 aactggggccaggagagagca 30  
||| |||||

Db 26429 AGCTCCGCCAGGAGGAGCA 26408

RESULT 13

US-08-457-342-6/c  
; Sequence 6, Application US/08457342  
; Patent No. 5662898  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James M.  
; APPLICANT: Beck, James Joseph  
; APPLICANT: Hill, Dwight Steven  
; APPLICANT: Ryals, John Andrew  
; APPLICANT: Gaffney, Thomas Deane  
; APPLICANT: Lam, Stephen Ting  
; APPLICANT: Hammer, Phillip E.  
; APPLICANT: Uknes, Scott Joseph  
; TITLE OF INVENTION: Genes for the synthesis of  
; TITLE OF INVENTION: antipathogenic substances  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ciba-Geigy Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/457,342  
; FILING DATE: 01-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/457,205  
; FILING DATE: 01-JUN-1995  
; APPLICATION NUMBER: 08/258,261  
; FILING DATE: 08-Jun-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Elmer, James Scott  
; REGISTRATION NUMBER: 36,129  
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8614  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28958 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-457-342-6

Query Match 57.3%; Score 17.2; DB 1; Length 28958;  
Best Local Similarity 86.4%; Pred. No. 97;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 9 aactggggccaggagagagca 30  
||| |||||

Db 26429 AGCTCCGCCAGGAGGAGCA 26408

RESULT 14

US-08-457-646A-6/c  
; Sequence 6, Application US/08457646A  
; Patent No. 5679560  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James M.  
; APPLICANT: Beck, James Joseph  
; APPLICANT: Hill, Dwight Steven  
; APPLICANT: Ryals, John Andrew  
; APPLICANT: Gaffney, Thomas Deane  
; APPLICANT: Lam, Stephen Ting  
; APPLICANT: Hammer, Phillip E.  
; APPLICANT: Uknes, Scott Joseph  
; TITLE OF INVENTION: Genes for the synthesis of  
; TITLE OF INVENTION: antipathogenic substances  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ciba-Geigy Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/457,646A  
; FILING DATE: 01-JUN-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/457,205  
; FILING DATE: 01-JUN-1995  
; APPLICATION NUMBER: 08/258,261  
; FILING DATE: 08-Jun-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Elmer, James Scott  
; REGISTRATION NUMBER: 36,129  
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8614  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 28958 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-457-646A-6

Query Match 57.3%; Score 17.2; DB 1; Length 28958;  
Best Local Similarity 86.4%; Pred. No. 97;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 9 aactggggccaggagagagca 30  
||| |||||

Db 26429 AGCTCCGCCAGGAGGAGCA 26408

Mon. May 20 10:14:54 2002

```

RESULT 15
US-08-458-076A-6/c
; Sequence 6, Application US/08458076A
; Patent No. 5698425
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,076A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-458-076A-6

```

```
Query Match      57.3%; Score 17.2; DB 1; Length 28958;
Best Local Similarity 86.4%; Pred. No. 97;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

Search completed: May 18, 2002, 17:26:47  
Job time: 12849 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:22:05 ; Search time 8624.33 Seconds  
(without alignments)  
31.300 Million cell updates/sec

Title: US-09-623-329-4

Perfect score: 20  
Sequence: 1 aatacagacaatggactccc 20

Scoring table: IDENTITY\_NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_esti:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	17	85.0	657	12 AG045201	AG045201 Pan trogl
c 2	17	85.0	688	12 AG126122	AG126122 Pan trogl
c 3	16.8	84.0	306	9 BB393281	BB393281 BB393281
c 4	16.8	84.0	397	9 AA009350	AA009350 TgEST2208
c 5	16.8	84.0	528	9 AW764851	AW764851 da92h05.x
c 6	16.8	84.0	561	12 AZ785010	AZ785010 2M0028C01
c 7	16.8	84.0	565	9 AW934950	AW934950 RC2-D7000
c 8	16.8	84.0	589	12 AZ782008	AZ782008 2M0021E23
c 9	16.8	84.0	649	10 BJ085382	BJ085382 BJ085382
c 10	16.8	84.0	678	9 BB030622	BB030622 BB030622
c 11	16.8	84.0	734	12 BH079030	BH079030 RPI-24-3
c 12	16.8	84.0	844	10 BG968060	BG968060 60282244
c 13	16.8	84.0	861	10 BF160622	BF160622 601768512
c 14	16.8	84.0	940	10 BF535641	BF535641 602054038
c 15	16.8	84.0	1091	11 AK008798	AK008798 Mus muscu
c 16	16.8	84.0	1095	12 CNS05P62	AL347555 Tetraodon
c 17	16.4	82.0	367	9 AI414308	AI414308 md85e04.x

18	16.4	82.0	371	9 BB871056	BB871056 BB871056
c 19	16.4	82.0	377	9 AI844195	AI844195 UI-M-AL1-
c 20	16.4	82.0	419	10 BE855078	BE855078 ux35a08.x
c 21	16.4	82.0	428	9 AI840102	AI840102 UI-M-AL0-
c 22	16.4	82.0	492	10 BF567756	BF567756 UI-R-B00-
c 23	16.4	82.0	510	12 A0785583	A0785583 HS_3077-A
c 24	16.4	82.0	558	10 BF286556	BF286556 EST451147
c 25	16.4	82.0	690	10 BE951512	BE951512 UI-M-CC0-
c 26	16.4	82.0	1101	12 CNS00117	AL061146 Drosophil
c 27	16	80.0	386	12 AQ840060	AQ840060 nbxb0051P
c 28	16	80.0	572	12 BH024612	BH024612 RPI-24-3
c 29	16	80.0	645	12 A0329291	A0329291 nbxb0044B
c 30	16	80.0	738	12 AG086523	AG086523 Pan trogl
c 31	16	80.0	1193	10 BE901969	BE901969 601676739
c 32	15.8	79.0	245	10 BM133458	BM133458 NXLV_006-
c 33	15.8	79.0	281	12 AQ934079	AQ934079 RPI-23-2
c 34	15.8	79.0	294	12 B33125	B33125 HS-1016-B1-
c 35	15.8	79.0	355	12 B82498	B82498 RPI11-16P8
c 36	15.8	79.0	382	10 BG317446	BG317446 NXPV_001-
c 37	15.8	79.0	408	10 BE239010	BE239010 MD0666 Me
c 38	15.8	79.0	420	12 AQ053328	AQ053328 CIT-HSP-2
c 39	15.8	79.0	427	9 AW936140	AW936140 QVO-D7002
c 40	15.8	79.0	432	9 AA532202	AA532202 TENF0140
c 41	15.8	79.0	436	9 AI504677	AI504677 vq34f02.x
c 42	15.8	79.0	438	10 BG673812	BG673812 NXPV_075-
c 43	15.8	79.0	441	12 TA278011P	AL485508 T. brucei
c 44	15.8	79.0	445	9 AA464162	AA464162 z883C10.S
c 45	15.8	79.0	451	12 AZ599371	AZ599371 IM0414K07

#### ALIGNMENT'S

RESULT 1  
AG045201/c

LOCUS Pan troglodytes DNA, clone: PTB-024A19.F, genomic survey sequence.  
DEFINITION AG045201  
ACCESSION AG045201  
VERSION AG045201.1 GI:16582018  
KEYWORDS GSS; GSS (genome survey sequence).  
SOURCE Pan troglodytes male lymphoblast DNA, clone\_lib:PTB Chimpanzee Male  
BAC Library clone:PTB-024A19.F.  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE  
AUTHORS

Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE  
JOURNAL

BAC end sequences of Library PTB  
Unpublished

REFERENCE  
AUTHORS

2 (bases 1 to 657)  
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE  
JOURNAL

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-chou,Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail:chimpbes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,  
Tel:81-45-503-9111, Fax:81-45-503-9170)

Clones are derived from the chimpanzee BAC library PTB This BAC end  
clone tracking errors.  
PRIMERS  
Sequencing: -21M13

LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1. .657

/organism="Pan troglodytes"

/db\_xref="taxon:9598"

/clone="PTB-024A19.F"

FEATURES  
source







```

KEYWORDS      GSS.
SOURCE         Mus musculus
ORGANISM       house mouse.

REFERENCE
AUTHORS       Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
              Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
              M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
              and Wright,D., Weiss,R.

TITLE         Mouse whole genome scaffolding with paired end reads from 10kb
              plasmid inserts

JOURNAL
COMMENT       Unpublished (2000)
              Contact: Robert B. Weiss
              University of Utah Genome Center
              University of Utah
              Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
              84112, USA
              Tel: 801 585 5606
              Fax: 801 585 7177
              Email: ddunndgenetics.utah.edu
              Insert Length: 10000 Std Error: 0.00
              Plate: 0028 row: C column: 01
              Seq primer: CACACAGGAACACGCTATGACC
              Class: plasmid ends
              High quality sequence stop: 561.

FEATURES      1..561
              Location/Qualifiers
              source
                1..561
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUGC2M0028C01"
                /clone_lib="Mouse 10kb plasmid UUGCIM library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, T1-resistant, F."
                /note="Vector: PWD42nv; Purified genomic DNA from M.
                musculus C57BL/6J (male) was obtained from the Jackson
                Laboratory Mouse DNA Resource
                (http://www.jax.org/resources/documents/dnares/). The DNA
                was hydrodynamically sheared by repeated passage through a
                0.005 inch orifice at constant velocity. The sheared DNA
                was blunt end-repaired with T4 DNA polymerase and T4
                polynucleotide kinase. Adaptor oligonucleotides were
                ligated to the blunt ends in high molar excess. The
                adaptor DNA was purified and size-selected for a 9.5 to
                10.5 kb range using preparative agarose gel
                electrophoresis. Vector DNA was prepared from a derivative
                of pMD42 (gi14732114|gb|AF129072.1), a copy-number
                inducible derivative of plasmid R1. The vector was ligated
                with adaptors complementary to the insert adaptors and
                purified. The sheared, adaptor mouse DNA was annealed to
                adaptor vector DNA, and transformed into
                chemically-competent E. coli XL10-Gold (Stratagene) cells
                and selected for ampicillin resistance."

BASE COUNT    215 a 90 c 95 g 160 t 1 others
ORIGIN

Query Match      84.0%; Score 16.8; DB 12; Length 561;
Best Local Similarity 90.0%; Pred. No. 6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aatacagacaatggactccc 20
    ||||| ||||| ||||| |||||
Db 215 AATAAAGACATGGATGCC 234

RESULT 7
LOCUS      AW934950
DEFINITION RC2-DR0002-261299-011-b01 DT0002 Homo sapiens cDNA, mRNA sequence.
ACCESSION  AW934950
VERSION     AW934950.1 GI:8110265

KEYWORDS      GSS.
SOURCE         Mus musculus
ORGANISM       house mouse.

REFERENCE
AUTHORS       Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
              Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
              M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
              and Wright,D., Weiss,R.

TITLE         Mouse whole genome scaffolding with paired end reads from 10kb
              plasmid inserts

JOURNAL
COMMENT       Unpublished (2000)
              Contact: Robert B. Weiss
              University of Utah Genome Center
              University of Utah
              Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
              84112, USA
              Tel: 801 585 5606
              Fax: 801 585 7177
              Email: ddunndgenetics.utah.edu
              Insert Length: 10000 Std Error: 0.00
              Plate: 0028 row: C column: 01
              Seq primer: CACACAGGAACACGCTATGACC
              Class: plasmid ends
              High quality sequence stop: 561.

FEATURES      1..561
              Location/Qualifiers
              source
                1..561
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUGC2M0028C01"
                /clone_lib="Mouse 10kb plasmid UUGCIM library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, T1-resistant, F."
                /note="Vector: PWD42nv; Purified genomic DNA from M.
                musculus C57BL/6J (male) was obtained from the Jackson
                Laboratory Mouse DNA Resource
                (http://www.jax.org/resources/documents/dnares/). The DNA
                was hydrodynamically sheared by repeated passage through a
                0.005 inch orifice at constant velocity. The sheared DNA
                was blunt end-repaired with T4 DNA polymerase and T4
                polynucleotide kinase. Adaptor oligonucleotides were
                ligated to the blunt ends in high molar excess. The
                adaptor DNA was purified and size-selected for a 9.5 to
                10.5 kb range using preparative agarose gel
                electrophoresis. Vector DNA was prepared from a derivative
                of pMD42 (gi14732114|gb|AF129072.1), a copy-number
                inducible derivative of plasmid R1. The vector was ligated
                with adaptors complementary to the insert adaptors and
                purified. The sheared, adaptor mouse DNA was annealed to
                adaptor vector DNA, and transformed into
                chemically-competent E. coli XL10-Gold (Stratagene) cells
                and selected for ampicillin resistance."

BASE COUNT    215 a 90 c 95 g 160 t 1 others
ORIGIN

Query Match      84.0%; Score 16.8; DB 12; Length 561;
Best Local Similarity 90.0%; Pred. No. 6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aatacagacaatggactccc 20
    ||||| ||||| ||||| |||||
Db 215 AATAAAGACATGGATGCC 234

RESULT 8
LOCUS      AZ782008/c
DEFINITION 2M0021E23R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
              clone UUGC2M0021E23 R, DNA sequence.
ACCESSION  AZ782008
VERSION     AZ782008.1 GI:12915271
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus.

REFERENCE
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
              Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
              M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
              and Wright,D., Weiss,R.

KEYWORDS      Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
              1 (bases 1 to 589)
              Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
              Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
              M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
              and Wright,D., Weiss,R.

```

```

KEYWORDS      EST.
SOURCE         human.
ORGANISM       Homo sapiens

REFERENCE
AUTHORS       Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
              Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
              Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
              Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
              ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
              Simpson,A.J.

TITLE         Shotgun sequencing of the human transcriptome with ORF expressed
              sequence tags

JOURNAL
COMMENT       Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
              20202663
              Contact: Simpson A.J.G.
              Laboratory of Cancer Genetics
              Ludwig Institute for Cancer Research
              Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
              Brazil
              Tel: +55-11-2704922
              Fax: +55-11-2707001
              Email: asimpson@ludwig.org.br
              This sequence was derived from the FAPES/ ICR Human Cancer Genome
              Project. This entry can be seen in the following URL
              (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-RC2-DT0002-261
              299-011-b01et3-1999-12-26et4-1)
              Seq primer: puc 18 forward
              High quality sequence stop: 565.

FEATURES      1..565
              Location/Qualifiers
              source
                1..565
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="DT0002"
                /dev_stage="Adult"
                /note="Organ: denis_drash; Vector: puc18; Site_1: Sma1;
                Site_2: Sma1; A mini-library was made by cloning products
                derived from ORESTES PCR (U.S. Letters Patent application
                No. 196,716 - Ludwig Institute for Cancer Research)
                profiles into the pUC 18 vector. Reverse transcription of
                tissue mRNA and cDNA amplification were performed under
                low stringency conditions."

BASE COUNT    136 a 105 c 146 g 178 t
ORIGIN

Query Match      84.0%; Score 16.8; DB 9; Length 565;
Best Local Similarity 90.0%; Pred. No. 6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aatacagacaatggactccc 20
    ||||| ||||| ||||| |||||
Db 194 AATACAGATAATGGATGCC 175

RESULT 8
LOCUS      AZ782008/c
DEFINITION 2M0021E23R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
              clone UUGC2M0021E23 R, DNA sequence.
ACCESSION  AZ782008
VERSION     AZ782008.1 GI:12915271
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus.

REFERENCE
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
              Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
              M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
              and Wright,D., Weiss,R.

```



sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

**Accession:** cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in KIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES  
SOURCE

```
Location/Qualifiers
1. 678
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="5830456f11"
/clone_lib="RIKEN f
/sex="male"
/tissue_type="thymus"
/dev_stage="adult"
/lab_host="DH10B"
```

prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTCGATTATTAATTAATTCCTCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pluciscript KS(+) after bulk excision from Lambda FIC I. Cloning sites, 5' end: SalI 3' end:

	BATH 1.		
BASE COUNT	151 a	169 c	176 g
ORIGIN			182 t

Query Match 84.0%; Score 16.8; DB 9; Length 678;  
Best Local Similarity 90.0%; Pred. No. 6.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aatacagacaatggactccc 20  
||| ||||||||| |||  
Db 385 AATTCAGACAAATGGACCCCC 404

Db 385 AATCAGACAATGGACCCC 404

RESULT 11

BH079030
LOCUS
DEFINITION
BH079030 734 bp DNA linear GSS 18-JUL-2001 RPCI-24-304G4.TJ RPCI-24 Mus musculus genomic clone RPCI-24-304G4, DNA sequence.

ACCESSION  
BH079030

VERSION BH079030.1 GI:14898627

**KEYWORDS**

SOURCE house mouse.  
ORGANISM *Mus musculus*

## ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE AUTHORS

TITLE	JOURNAL	COMMENT
1. The Role of the Teacher in the Classroom	Journal of Educational Research	1965, Vol. 68, No. 1, pp. 1-10
2. The Impact of Technology on Education	Journal of Educational Technology	1978, Vol. 1, No. 2, pp. 1-10
3. The Importance of Parental Involvement	Journal of Educational Research	1985, Vol. 88, No. 3, pp. 1-10
4. The Effect of Teacher Expectations on Student Achievement	Journal of Educational Research	1990, Vol. 93, No. 4, pp. 1-10
5. The Role of the School in the Community	Journal of Educational Research	1995, Vol. 98, No. 5, pp. 1-10
6. The Impact of Teacher Education on Student Achievement	Journal of Educational Research	2000, Vol. 103, No. 6, pp. 1-10
7. The Role of the Teacher in the 21st Century	Journal of Educational Research	2005, Vol. 108, No. 7, pp. 1-10
8. The Impact of Teacher Professional Development on Student Achievement	Journal of Educational Research	2010, Vol. 113, No. 8, pp. 1-10
9. The Role of the Teacher in the Global Classroom	Journal of Educational Research	2015, Vol. 118, No. 9, pp. 1-10
10. The Impact of Teacher Leadership on Student Achievement	Journal of Educational Research	2020, Vol. 123, No. 10, pp. 1-10

FEATURES	SOURCE
1. <b>High Accuracy:</b> The model achieves a high accuracy rate, indicating its effectiveness in predicting the target variable.	Model Performance Metrics
2. <b>Robustness:</b> The model demonstrates robustness against various input variations and noise, ensuring reliable predictions.	Model Performance Metrics
3. <b>Scalability:</b> The model is designed to handle large-scale data efficiently, making it suitable for big data applications.	Model Performance Metrics
4. <b>Interpretability:</b> The model's predictions are interpretable, allowing users to understand the underlying factors influencing the results.	Model Performance Metrics
5. <b>Real-time Processing:</b> The model is optimized for real-time data processing, enabling immediate insights and decision-making.	Model Performance Metrics
6. <b>Customizable:</b> The model can be tailored to specific use cases and requirements, providing flexibility in its application.	Model Performance Metrics
7. <b>Integration:</b> The model seamlessly integrates with existing systems and workflows, facilitating easy adoption.	Model Performance Metrics
8. <b>Security:</b> The model adheres to strict security protocols, ensuring that data and predictions are protected from unauthorized access.	Model Performance Metrics
9. <b>Compliance:</b> The model is designed to comply with relevant industry regulations and standards, ensuring legal and ethical use.	Model Performance Metrics
10. <b>Support:</b> Comprehensive support and documentation are provided to assist users in maximizing the model's potential.	Model Performance Metrics

```

FEATURES
  source
    Location/Qualifiers
      i. 734
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="RPCI-24-304G4"
        /clone_lib="RPCI-24"
        /sex="Male"
        /cell_type="Spleen/Brain"
        /note="Vector: pTARBAC1; Site
RPCI-24 Mouse HAC Library p
library was cloned in the
BamHI sites using MboI par

```

	Data		
BASE COUNT	256 a	144 c	129 g
ORIGIN			205 t

Query Match  
Best Local  
Matches

Qy 1 aatacagacaatggactccc 20

Db 637 AATAAGACAAGGACTCCC 656

RESULT 12  
RG96806070

SN:001  
Locus

### DEFINITION

mRNA sequ

ACCESSION	VERSION	CT
BG968060	BC069060	1
		CT.14255607

VERSION	BG96
KEYWORDS	FST

KEYWORDS	EST.
SOURCE	house mouse.

**SOURCE**  
**ORGANISM**

55

100

## REFERENCE

AUTHORS	
TITLE	

TITLE  
JOURNAL.





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:26:27 ; Search time 250.21 Seconds  
(without alignments)  
19.634 Million cell updates/sec

Title: US-09-623-329-4

Perfect score: 20

Sequence: 1 aatacagacaatggactccc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued\_Patents\_NA:\*
- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*
  - 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*
  - 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*
  - 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*
  - 5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*
  - 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	17	85.0	1748	3	US-09-100-730-1
c 2	17	85.0	5452	2	US-09-130-114-1
c 3	17	85.0	8575	5	PCT-US92-08258-6
c 4	17	85.0	9600	4	US-08-910-647-1
c 5	17	85.0	10596	1	US-07-884-811-15
c 6	17	85.0	10596	1	US-07-885-971-15
c 7	17	85.0	10596	1	US-08-087-783A-15
c 8	17	85.0	10596	1	US-08-194-088B-15
c 9	17	85.0	10596	2	US-08-194-087-15
c 10	17	85.0	10596	5	PCT-US93-04648-15
c 11	16.8	84.0	889	4	US-09-227-357-88
c 12	16	80.0	2580	3	US-09-050-863-2
c 13	16	80.0	2580	4	US-09-359-081-2
c 14	15.2	76.0	280	3	US-09-026-343-31
c 15	15	75.0	9793	1	US-08-470-202-56
c 16	15	75.0	9793	2	US-08-471-770-56
c 17	15	75.0	9793	1	US-08-468-059-56
c 18	15	75.0	9793	4	US-09-109-916-56
c 19	14.8	74.0	1836	1	US-07-929-580B-1
c 20	14.8	74.0	1839	1	US-07-929-580B-4
c 21	14.4	72.0	2911	3	US-08-981-825-5
c 22	14.4	72.0	2911	4	US-09-480-784-5
c 23	14.4	72.0	4260	2	US-08-658-665-38
c 24	14.4	72.0	4260	4	US-08-796-101-2
c 25	14.4	72.0	4260	4	US-09-085-273-38
c 26	14.4	72.0	5234	2	US-08-658-665-73
c 27	14.4	72.0	5234	4	US-08-796-101-37

c 28	14.4	72.0	5234	4	US-09-085-273-73	Sequence 73, Appl
c 29	14.4	72.0	6601	1	US-08-232-463-15	Sequence 15, Appl
c 30	14.4	72.0	7218	1	US-08-232-463-14	Sequence 14, Appl
c 31	14.4	72.0	7252	4	US-09-238-356-27	Sequence 27, Appl
c 32	14.4	72.0	8710	1	US-08-480-882B-3	Sequence 3, Appl
c 33	14.4	72.0	8710	1	US-08-480-210-3	Sequence 3, Appl
c 34	14.4	72.0	9019	1	US-08-480-882B-4	Sequence 4, Appl
c 35	14.4	72.0	9019	1	US-08-480-210-4	Sequence 4, Appl
c 36	14.4	72.0	9890	1	US-08-232-463-18	Sequence 18, Appl
c 37	14.4	72.0	9916	1	US-08-232-463-17	Sequence 17, Appl
c 38	14.4	72.0	9917	1	US-08-232-463-16	Sequence 16, Appl
c 39	14.4	72.0	10408	1	US-08-232-463-6	Sequence 6, Appl
c 40	14.4	72.0	10408	1	US-08-232-463-7	Sequence 7, Appl
c 41	14.4	72.0	13910	4	US-09-263-933-1	Sequence 1, Appl
c 42	14.4	72.0	13910	4	US-09-263-933-8	Sequence 8, Appl
c 43	14.4	72.0	13910	4	US-09-263-933-15	Sequence 15, Appl
c 44	14.2	71.0	81	4	US-09-133-321-7	Sequence 7, Appl
c 45	14.2	71.0	399	4	US-09-328-111-352	Sequence 352, App

ALIGNMENTS

RESULT 1  
US-09-100-730-1/c  
; Sequence 1, Application US/09100730  
; Patent No. 6133025  
; GENERAL INFORMATION:  
; APPLICANT: Seed, Brian  
; TITLE OF INVENTION: COMPACT EPSTEIN-BARR VIRUS REPLICONS  
; FILE REFERENCE: 00786/357001  
; CURRENT APPLICATION NUMBER: US/09/100,730  
; CURRENT FILING DATE: 1998-06-19  
; EARLIER APPLICATION NUMBER: 60/050,206  
; EARLIER FILING DATE: 1997-06-19  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 1748  
; TYPE: DNA  
; ORGANISM: Epstein Barr Virus  
US-09-100-730-1

Query Match 85.0%; Score 17; DB 3; Length 1748;  
Best Local Similarity 100.0%; Pred. No. 3.3;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 acatgacatggactccc 20  
Db 1475 ACAGACAATGGACTCCC 1459  
|||||

RESULT 2  
US-09-130-114-1  
; Sequence 1, Application US/09130114  
; Patent No. 5978607  
; GENERAL INFORMATION:  
; APPLICANT: Horlick, Robert A.  
; APPLICANT: Damaj, Bassam B.  
; APPLICANT: Robbins, Alan K.  
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes  
; FILE REFERENCE: 0867/1D903051  
; CURRENT APPLICATION NUMBER: US/09/130,114  
; CURRENT FILING DATE: 1998-08-06  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 5452  
; TYPE: DNA  
; ORGANISM: VEBNA  
US-09-130-114-1

Query Match 85.0%; Score 17; DB 2; Length 5452;  
Best Local Similarity 100.0%; Pred. No. 4;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 acagacaatggactccc 20  
Db 742 acagacaatggactccc 758

## RESULT 3

PCT-US92-08258-6/c  
; Sequence 6, Application PC/TUS9208258  
; GENERAL INFORMATION:  
; APPLICANT: GENENTECH, INC. and STATE OF OREGON BY AND  
; TITLE OF INVENTION: PRODUCTION OF GPA NEUTROTROPIC FACTOR  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)  
; CURRENT APPLICATION DATA:  
; FILING DATE: 19920929  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Johnston, Sean A.  
; REGISTRATION NUMBER: P35,910  
; REFERENCE/DOCKET NUMBER: 731  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-3562  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8575 bases  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
PCT-US92-08258-6

Query Match 85.0%; Score 17; DB 5; Length 8575;  
Best Local Similarity 100.0%; Pred. No. 4.4;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 acagacaatggactccc 20  
Db 234 ACAGACAATGGACTCCC 218

## RESULT 4

US-08-910-647-1/c  
; Sequence 1, Application US/08910647  
; Patent No. 6251433  
; GENERAL INFORMATION:  
; APPLICANT: Zuckermann et al.  
; TITLE OF INVENTION: Compositions and Methods for  
; POLYNUCLEOTIDE DELIVERY  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Chiron Corporation  
STREET: 4560 Horton Street  
CITY: Emeryville  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 94608-2916  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/910.647  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Fujita, Sharon M.  
REGISTRATION NUMBER: 38,459  
REFERENCE/DOCKET NUMBER: 1218.002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 923-2706  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9600 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-910-647-1

Query Match 85.0%; Score 17; DB 4; Length 9600;  
Best Local Similarity 100.0%; Pred. No. 4.4;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 acagacaatggactccc 20

Db 2109 ACAGACAATGGACTCCC 2093

## RESULT 5

US-07-884-811-15/G  
; Sequence 15, Application US/07884811  
; Patent No. 5316921  
; GENERAL INFORMATION:  
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.  
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/884,811  
; FILING DATE: 19920518  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Ginger R.  
; REGISTRATION NUMBER: 33,055  
; REFERENCE/DOCKET NUMBER: 755.1  
; TELECOMMUNICATION INFORMATION:



```

; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-07-884-811-15

Query Match 85.0%; Score 17; DB 1; Length 10596;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 acagacaatgactccc 20
Db 3643 ACAGACAATGACTCCC 3627

RESULT 6
US-07-885-971-15/c
; Sequence 15, Application US/07885971
; Patent No. 532837
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,971
; FILING DATE: 19920518
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; ATTORNEY/AGENT INFORMATION:
; FILING DATE:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-07-885-971-15

Query Match 85.0%; Score 17; DB 1; Length 10596;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 acagacaatgactccc 20
Db 3643 ACAGACAATGACTCCC 3627

RESULT 7
US-08-087-783A-15/c
; Sequence 15, Application US/08087783A
; Patent No. 5547856
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatlin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/087,783A
; FILING DATE: 13-Jul-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/884811
; FILING DATE: 18-MAY-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/885971
; FILING DATE: 18-MAY-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Marschang, Diane L.
; REGISTRATION NUMBER: 35,600
; REFERENCE/DOCKET NUMBER: P0755779P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5416
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
;
US-08-087-783A-15

Query Match 85.0%; Score 17; DB 1; Length 10596;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 acagacaatgactccc 20
Db 3643 ACAGACAATGACTCCC 3627

RESULT 8
US-08-194-088B-15/c
; Sequence 15, Application US/08194088B
; Patent No. 5580963
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
```

;; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: patin (Genentech)  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/194,088B  
;; FILING DATE: 09-FEB-1994  
;; CLASSIFICATION: 530  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 07/884811  
;; FILING DATE: 18-MAY-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Gallegos, R. Thomas  
;; REGISTRATION NUMBER: 32,692  
;; REFERENCE/DOCKET NUMBER: 75501  
;; TELEPHONE: 415/225-2614  
;; TELEFAX: 415/952-9881  
;; TELEX: 910/371-7168  
;; INFORMATION FOR SEQ ID NO: 15:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 10596 bases  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; US-08-194-088B-15

Query Match 85.0%; Score 17; DB 1; Length 10596;  
Best Local Similarity 100.0%; Pred. No. 4.5;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 acagacaaatggactccc 20  
|||||

Db 3643 ACAGACAATGGACTCCC 3627

## RESULT 9

;; US-08-194-087-15/c  
;; Sequence 15, Application US/08194087  
;; Patent No. 5879910  
;; GENERAL INFORMATION:  
;; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.  
;; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS  
;; NUMBER OF SEQUENCES: 21  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Genentech, Inc.  
;; STREET: 460 Point San Bruno Blvd  
;; CITY: South San Francisco  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 94080  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patin (Genentech)  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/194,087  
;; FILING DATE: 18-MAY-1992  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER:  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Dreger, Ginger R.  
;; REGISTRATION NUMBER: 33,055  
;; REFERENCE/DOCKET NUMBER: 779  
;; TELEPHONE: 415/225-3216  
;; TELEFAX: 415/952-9881  
;; TELEX: 910/371-7168  
;; INFORMATION FOR SEQ ID NO: 15:

;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 10596 bases  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; US-08-194-087-15

Query Match 85.0%; Score 17; DB 2; Length 10596;  
Best Local Similarity 100.0%; Pred. No. 4.5;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 acagacaaatggactccc 20  
|||||

Db 3643 ACAGACAATGGACTCCC 3627

## RESULT 10

;; PCT-US93-04648-15/c  
;; Sequence 15, Application PC/TUS9304648  
;; GENERAL INFORMATION:  
;; APPLICANT: Genentech, Inc., Godowski, Paul J., Lokker, Natalie A., Mark, Melani  
;; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS  
;; NUMBER OF SEQUENCES: 21  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Genentech, Inc.  
;; STREET: 460 Point San Bruno Blvd  
;; CITY: South San Francisco  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 94080  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: patin (Genentech)  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US93/04648  
;; FILING DATE: 19930517  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 07/884811  
;; FILING DATE: 18-MAY-92  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 07/885971  
;; FILING DATE: 18-MAY-92  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Dreger, Ginger R.  
;; REGISTRATION NUMBER: 33,055  
;; REFERENCE/DOCKET NUMBER: 755,779p1  
;; TELEPHONE: 415/225-3216  
;; TELEFAX: 415/952-9881  
;; TELEX: 910/371-7168  
;; INFORMATION FOR SEQ ID NO: 15:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 10596 bases  
;; TYPE: NUCLEIC ACID  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; PCT-US93-04648-15

Query Match 85.0%; Score 17; DB 5; Length 10596;  
Best Local Similarity 100.0%; Pred. No. 4.5;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 acagacaaatggactccc 20  
|||||

Db 3643 ACAGACAATGGACTCCC 3627

## RESULT 11

US-09-227-357-88/c  
; Sequence 88, Application US/09227357  
; Patent No. 6342581  
; GENERAL INFORMATION:  
; APPLICANT: Fischer et al.  
; TITLE OF INVENTION: 123 Human Secreted Proteins  
; FILE REFERENCE: P2010F1  
; CURRENT APPLICATION NUMBER: US/09/227,357  
; CURRENT FILING DATE: 1999-01-08  
; EARLIER APPLICATION NUMBER: PCT/US98/13684  
; EARLIER FILING DATE: 1998-07-07  
; EARLIER APPLICATION NUMBER: 60/051,926  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,793  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,925  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,929  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,803  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,732  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,931  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,932  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,916  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,930  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,918  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,920  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,733  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,795  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,919  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,928  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/055,722  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,723  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,948  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,949  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,953  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,950  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,947  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,964  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/056,360  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,684  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,984  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/055,954  
; EARLIER FILING DATE: 1997-08-18  
; EARLIER APPLICATION NUMBER: 60/058,785  
; EARLIER FILING DATE: 1997-09-12  
; EARLIER APPLICATION NUMBER: 60/058,664  
; EARLIER FILING DATE: 1997-09-12  
; EARLIER APPLICATION NUMBER: 60/058,660  
; EARLIER FILING DATE: 1997-09-12

; EARLIER APPLICATION NUMBER: 60/058,661  
; EARLIER FILING DATE: 1997-09-12  
; NUMBER OF SEQ ID NOS: 672  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 88  
; LENGTH: 889  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (117)  
; OTHER INFORMATION: n equals a,t,g, or c  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (292)  
; OTHER INFORMATION: n equals a,t,g, or c  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (341)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-227-357-88  
  
Query Match 84.0%; Score 16.8; DB 4; Length 889;  
Best Local Similarity 90.0%; Pred. No. 3.8;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 aatacagacaatggactccc 20  
||||||| ||||| |||||  
DB 178 AATACAGATAATGGAATCCC 159  
  
RESULT 12  
US-09-050-863-2/c  
; Sequence 2, Application US/09050863  
; Patent No. 6114111  
; GENERAL INFORMATION:  
; APPLICANT: Lao, Ying  
; APPLICANT: Hiang, Betty  
; APPLICANT: Payan, Don  
; TITLE OF INVENTION: Mammalian Protein Interaction Cloning  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/050,863  
; FILING DATE: 30-MAR-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silva, Robin M.  
; REGISTRATION NUMBER: 38,304  
; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 949-8711  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2580 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA

US-09-050-863-2

Query Match 80.0%; Score 16; DB 3; Length 2580;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 acagacaatgactcc 19  
|||||

Db 2062 ACAGACAATGACTCC 2047

RESULT 13

US-09-359-081-2/c

: Sequence 2, Application US/09359081

: Patent No. 6316223

: GENERAL INFORMATION:

: APPLICANT: Lao, Ying

: Hiang, Betty

: Payan, Don

: TITLE OF INVENTION: Mammalian Protein Interaction Cloning

: System

: NUMBER OF SEQUENCES: 5

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Fiehr, Hobbach, Test, Albritton &amp; Herbert

: STREET: 4 Embarcadero Center, Suite 3400

: CITY: San Francisco

: STATE: CA

: COUNTRY: USA

: ZIP: 94111-4187

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: PatentIn Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/09/359,081

: FILING DATE: 22-Jul-1999

: CLASSIFICATION: &lt;Unknown&gt;

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: 09/050,863

: FILING DATE: &lt;Unknown&gt;

: ATTORNEY/AGENT INFORMATION:

: NAME: Silva, Robin M.

: REGISTRATION NUMBER: 38,304

: REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (415) 781-1989

: TELEFAX: (415) 949-8711

: INFORMATION FOR SEQ ID NO: 2:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 2580 base pairs

: TYPE: nucleic acid

: STRANDEDNESS: unknown

: TOPOLOGY: unknown

: MOLECULE TYPE: DNA

: SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-359-081-2

Query Match 80.0%; Score 16; DB 4; Length 2580;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 acagacaatgactcc 19  
|||||

Db 2062 ACAGACAATGACTCC 2047

RESULT 14

US-09-026-343-31/c

: Sequence 31, Application US/09026343

: Patent No. 6008018

: GENERAL INFORMATION:

: APPLICANT: DUAN, D. ROXANNE

: APPLICANT: SHILATIPARD, ALI

: APPLICANT: CONAWAY, JOAN W.

: APPLICANT: CONAWAY, RONALD C.

: TITLE OF INVENTION: ELL2, A New Member of an ELL Family of

: RNA Polymerase II Elongation Factors

: NUMBER OF SEQUENCES: 34

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: STERNE, KESSLER, GOLDSTEIN &amp; FOX P.L.L.C.

: STREET: 1100 NEW YORK AVENUE, SUITE 600

: CITY: WASHINGTON

: STATE: D.C.

: COUNTRY: USA

: ZIP: 20005-3934

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: PatentIn Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/09/026,343

: FILING DATE: Herewith

: CLASSIFICATION:

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: US 60/038,447

: FILING DATE: 19-FEB-1997

: ATTORNEY/AGENT INFORMATION:

: NAME: GOLDSTEIN, JORGE A.

: REGISTRATION NUMBER: 29,021

: REFERENCE/DOCKET NUMBER: 1488.08.001

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (202) 371-2600

: TELEFAX: (202) 372-2540

: INFORMATION FOR SEQ ID NO: 31:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 280 base pairs

: TYPE: nucleic acid

: STRANDEDNESS: both

: TOPOLOGY: both

: MOLECULE TYPE: cDNA

US-09-026-343-31

Query Match 76.0%; Score 15.2; DB 3; Length 280;  
Best Local Similarity 85.0%; Pred. No. 22;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 aatacagacaatgactccc 20  
|||||

Db 274 AATATAGACAATGATCCCC 255  
|||||

RESULT 15

US-08-470-202-56

: Sequence 56, Application US/08470202

: Patent No. 5759808

: GENERAL INFORMATION:

: APPLICANT: Guertler, Lutz G.

: APPLICANT: Eberle, Josef

: APPLICANT: Brunn, Albrecht V.

: APPLICANT: Knapp, Stefan

: APPLICANT: Hauser, Hans-Peter

: TITLE OF INVENTION: Retrovirus from the HIV Group and Its

: NUMBER OF SEQUENCES: 63

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &amp;

: STREET: 1300 I Street, N.W.

: CITY: Washington

: STATE: D.C.

: COUNTRY: USA

21P: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/470,202  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/132,653  
FILING DATE: 05-OCT-1993  
APPLICATION NUMBER: DE P 42 33 646.5  
FILING DATE: 06-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 42 35 718.7  
FILING DATE: 22-OCT-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 42 44 541.8  
FILING DATE: 30-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 43 18 186.4  
FILING DATE: 01-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Michael J. Blake  
REGISTRATION NUMBER: 37,096  
REFERENCE/DOCKET NUMBER: 05495-0001-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 56:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9793 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-470-202-56

Query Match 75.0%; Score 15; DB 1; Length 9793;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 atacagacaatggac 16  
Db 4600 ATACAGACAATGGAC 4614  
|||||

Search completed: May 18, 2002, 17:26:29  
Job time: 12831 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 17:45:51 ; Search time 1024.22 Seconds  
(without alignments)  
33.526 Million cell updates/sec

Title: US-09-623-329-4

Perfect score: 20

Sequence: 1 aatacagacaatggactccc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_032802:\*

1: /SID55/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*

2: /SID55/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*

3: /SID55/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*

4: /SID55/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*

5: /SID55/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*

6: /SID55/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*

7: /SID55/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*

8: /SID55/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*

9: /SID55/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*

10: /SID55/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*

11: /SID55/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*

12: /SID55/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*

13: /SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*

14: /SID55/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*

15: /SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*

16: /SID55/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*

17: /SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*

18: /SID55/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*

19: /SID55/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*

20: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*

21: /SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*

22: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*

23: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*

24: /SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	AAZ11654	Oligo specific for primer 483R. Unid
2	17	85.0	26	AAC88342	PCR-generated eEBN
3	17	85.0	624	AAT15397	Epstein-Barr nucle
4	17	85.0	1092	AA03478	EBV nuclear antigen
5	17	85.0	1212	AAT15396	Epstein-Barr virus
6	17	85.0	1748	AA03475	Epstein-Barr virus
7	17	85.0	1925	AA030924	Epstein-Barr virus
8	17	85.0	1926	AAA50254	Epstein-Barr virus
9	17	85.0	1926	AAF82902	EBV tethering prot

10	17	85.0	5452	20	AA090923	Anti-sense strand
11	17	85.0	8575	14	AA040324	Vector pSHuttle DN
12	17	85.0	8705	20	AA223778	Vector plasmid pCM
13	17	85.0	9600	19	AAV21683	Expression vector
14	17	85.0	9725	21	AA053873	Expression vector
15	17	85.0	9732	21	AA053879	Expression vector
16	17	85.0	9738	21	AA053874	Expression vector
17	17	85.0	9873	21	AA053875	Expression vector
18	17	85.0	10054	21	AA053876	Expression vector
19	17	85.0	10380	20	AA222248	Nucleotide sequenc
20	17	85.0	10596	14	AA051731	Plasmid pCisEBON f
21	17	85.0	10596	17	AA040348	Plasmid pCisEBON f
22	17	85.0	10596	20	AA015650	Nucleotide sequenc
23	17	85.0	11265	19	AAV59501	Plasmid pREP7::CTL
24	17	85.0	16080	21	AA059553	DNA clone pCEK Cl.
25	16.8	84.0	397	22	AA06284	Novel human polynu
26	16.8	84.0	798	20	AA228290	Rat neuronal limed
27	16.8	84.0	889	20	AA027388	Human secreted pro
28	16.8	84.0	10877	21	AA255620	Human BMP-7 gene 5
29	16	80.0	2580	21	AA075454	Nucleotide sequenc
30	15.8	79.0	32768	20	AA020515	Polynucleotide seq
31	15.4	77.0	51	22	AA028083	Human SNP oligonuc
32	15.4	77.0	852	20	AA084436	Human secreted pro
33	15.4	77.0	912	22	AA083219	Human secreted pro
34	15.4	77.0	1463	22	AA033488	Human colon cancer
35	15.4	77.0	1583	22	AA098240	Murine EST-derived
36	15.4	77.0	1950	22	AA044256	Human breast cell
37	15.4	77.0	1950	22	ABA54706	Human foetal liver
38	15.4	77.0	1950	22	ABA44488	Probe #2954 for ge
39	15.4	77.0	1950	22	AA02990	Human bone marrow
40	15.4	77.0	1950	22	AA028438	Probe #2939 for ge
41	15.4	77.0	1950	22	AA013006	Probe #3047 used t
42	15.4	77.0	1950	22	AA013461	Probe #2913 used t
43	15.4	77.0	1950	22	AA02922	Human hydrophobic
44	15.4	77.0	2001	22	AA094442	Human cDNA sequenc
45	15.4	77.0	2345	22	AA015842	

## ALIGNMENTS

### RESULT 1

AAZ11654  
ID AAZ11654 standard; DNA; 20 BP.

AC AAZ11654;

DT 19-NOV-1999 (first entry)

DE Oligo specific for Epstein Barr Nuclear Antigen 1 (EBNA-1) RNA.

KW Epstein Barr Virus; EBV infection; viral; gene transcription; EBNA-1;

KW Epstein Barr early RNA; Epstein Barr nuclear antigen 1; EBNA-1; LMP-1;

KW latent membrane protein; LMP-2; vIL10; BCRF-1; BARF1; BDLF2; NASBA;

KW EBV-associated malignancy; primer; ss.

OS Synthetic.

OS Epstein-Barr virus.

PN WO9945155-A2.

PD 10-SEP-1999.

PF 01-MAR-1999; 99WO-EP01392.

PR 04-MAR-1998; 98EP-0200655.

PR 14-DEC-1998; 98EP-0204231.

PA (ALKU ) AKZO NOBEL NV.

PI Vervoort MBHJ, Van Den Brule AJC, Middelorp JM;

XX WPI; 1999-551051/46.

XX Identifying Epstein Barr Virus infection -  
 XX Claim 11; Page 20; 50pp; English.  
 XX The invention provides methods for identifying an Epstein Barr Virus  
 CC (EBV) infection, that comprises determining viral gene transcription  
 CC patterns by amplification of specific RNA sequences. The binding sites  
 CC of the oligos suitable for amplification are located in the following  
 CC genes: Epstein Barr early RNA (EBER-1), Epstein Barr nuclear antigen 1  
 CC (EBNA-1), latent membrane protein 1 (LMP-1), LMP-2, and vIL10 (BCRF-1),  
 CC BART1 and BDLF2. The method comprises (a) amplifying a target sequence  
 CC within one or more RNA(s) transcribed from above gene sequences and the  
 CC (b) detecting the amplified products, determining the transcription  
 CC pattern and identifying the corresponding EBV-associated malignancy. The  
 CC RNA is amplified using a transcription based amplification technique  
 CC such as NASBA. The invention is used to diagnose malignant and  
 CC non-malignant EBV-associated diseases. Sequences AAZ11651-54 represent  
 CC oligos specific for EBNA-1 RNA.  
 XX Sequence 20 BP; 8 A; 6 C; 3 G; 3 T; 0 other;

Query Match 100.0%; Score 20; DB 20; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 0.43;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 aatacagacaatggactccc 20  
 Db 1 aatacagacaatggactccc 20

RESULT 2  
 AAC88342  
 (ID AAC88342 standard; DNA; 26 BP.  
 XX AAC88342;  
 XX 02-MAR-2001 (first entry)  
 DT primer 483R.  
 DE  
 XX Nasopharyngeal carcinoma; Epstein Barr virus; screening; ss.

XX Unidentified.  
 XX WO200066769-A2.  
 XX 09-NOV-2000.  
 XX 28-APR-2000; 2000WO-CA00456.  
 XX 30-APR-1999; 99US-0131944.  
 XX (ADSE-) ADVANCE SENTRY CORP.  
 XX Ng RHW, Daykin V, Phillips J;  
 XX WPI; 2001-007233/01.  
 XX Screening nasopharyngeal carcinoma comprises quantifying the amount of  
 PT cellular Epstein Barr virus in control and test samples to define  
 PT threshold and test values, respectively, which are then compared -  
 XX Claim 6; Page 18; 36pp; English.

XX The present invention relates to screening nasopharyngeal carcinoma  
 CC and involves quantifying an amount of cellular Epstein Barr virus in  
 CC epithelial cell samples from nasopharynx of control and test patients  
 CC to define a threshold and test value.  
 XX Sequence 26 BP; 10 A; 7 C; 4 G; 5 T; 0 other;

Query Match 85.0%; Score 17; DB 22; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 acagacaatggactccc 20  
 Db 5 acagacaatggactccc 21

RESULT 3  
 AAT15397/C  
 ID AAT15397 standard; DNA; 624 BP.

XX AAT15397;  
 AC  
 XX 19-JUL-1996 (first entry)  
 DT PCR-generated eEBNA1 protein gene.  
 XX EBV; nuclear antigen protein 1; EBNA1; immunoassay; ds.  
 XX Epstein-barr virus.

XX Key Location/Qualifiers  
 FT primer\_bind complement (1..24)  
 FT /\*tag= a  
 FT /note= "primer for eEBNA1"  
 FT primer\_bind 584..624  
 FT /\*tag= b  
 FT /note= "primer for eEBNA1"

XX WO9602563-A1.  
 XX 01-FEB-1996.  
 XX 13-JUL-1995; 95WO-US08700.  
 XX 13-JUL-1994; 94US-0275614.  
 XX (CORR ) CORNELL RES FOUND INC.

XX O'Donnell ME;  
 XX WPI; 1996-105853/11.  
 XX P-PSDB; AAR88588.  
 XX Recovery of Epstein Barr Virus nuclear antigen protein 1 (EBNA1) -  
 PT useful in immunoassay system for detecting Epstein Barr Virus in  
 PT serum samples

XX Example 22; Page 51; 82pp; English.

XX A DNA fragment (AAT15397) was generated by PCR amplification of  
 CC the Epstein-Barr virus (EBV) nuclear antigen protein 1 (EBNA1)  
 CC gene (see also AAT15396) in plasmid p291 and NdeI/BamHI digestion of  
 CC the PCR product. It was ligated into pET3c and the resulting  
 CC plasmid was used to transform Escherichia coli strain BL21  
 CC (DE3)pLYSS cells. The recombinant EBNA1, designated eEBNA1  
 CC (AAR88588), can be recovered from the nuclei of host cells for use  
 CC EBV immunoassays.

XX Sequence 624 BP; 148 A; 132 C; 195 G; 149 T; 0 other;

Query Match 85.0%; Score 17; DB 17; Length 624;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 acagacaatggactccc 20  
 Db 357 ACAGACAATGGACTCCC 341





```

FT      /*tag= b
FT      /product= Epstein-Barr nuclear antigen 1
FT      /note= "EBNA1, specifically claimed in Claim 4"
XX
XX      WO9857658-A1.
XX
XX      23-DEC-1998.
XX
XX      14-APR-1998; 98WO-US07641.
XX
XX      19-JUN-1997; 97US-0050206.
XX
XX      (GEHO ) GEN HOSPITAL CORP.
XX
XX      Seed B;
XX
XX      WPI; 1999-080933/07.
XX
XX      Epstein-Barr virus replicons which support episomal replication
XX      comprise an Orip sequence and an EBNA1 sequence operably linked to a
XX      promoter
XX
XX      Disclosure; Fig 1; 12pp; English.
XX
XX      This 1748 bp fragment acts as a compact Epstein-Barr virus (EBV)
XX      replicon. Extensive deletion and mutagenesis studies showed that
XX      it is possible to embed the cis- and trans-acting functions
XX      necessary for EBV episomal replication into a fragment of less than
XX      2 kb. The 1748 bp fragment contains all sequences needed for
XX      efficient expression of the Epstein-Barr nuclear antigen 1 (EBNA-1)
XX      protein, with the exception of a polyA consensus sequence. It is
XX      designed to be inserted upstream of a bidirectional polyA sequence
XX      (see also AAX03476) in an appropriate vector. Plasmid vectors based
XX      on this sequence replicate as episomes in the nucleus of transfected
XX      mammalian cells. Compact EBV replicons find use in gene therapy
XX      vectors, e.g. in gene delivery vehicles such as expression vectors.
XX
XX      Sequence 1748 BP; 453 A; 382 C; 546 G; 367 T; 0 other;
SQ

```

```

Query Match      85.0%; Score 17; DB 20; Length 1748;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      4 acagacaatggactccc 20
        |||
Db      1475 ACAGACAATGGACTCCC 1459

```

```

RESULT 7
AAX90924/C
ID      AAX90924 standard; DNA; 1925 BP.
XX
XX      AAX90924;
XX
XX      17-JAN-2000 (first entry)
XX
XX      Epstein Barr Virus Nuclear Antigen 1 (EBNA 1) DNA.
XX
XX      Epstein Barr Virus Nuclear Antigen 1 DNA; plasmid pCMVEBNA; EBNA 1;
XX      episome; transfection; origin of replication; EBV orip; receptor;
XX      eucaryotic host cell; recombinant cell line; ion channel; gene therapy;
XX      multiple gene expression; transporter protein; transcription factor;
XX      adhesion molecule; antisense therapy; gene amplification;
XX      cell immortalisation; ds.
XX
XX      Epstein-barr virus.
XX
XX      Key      Location/Qualifiers
XX      CDS      1..1925
XX      FT      /*tag= a
XX      FT      /product= "EBNA 1"
XX      FT      /transl_except= (pos:799..800, aa:Gly)
XX

```

```

FT      /note= "The sequence is described throughout the
FT      specification as being 1926 nucleotides long, but a
FT      sequence of only 1925 bp has been given in figure 2"
XX
XX      WO9947647-A1.
XX
XX      23-SEP-1999.
XX
XX      12-FEB-1999; 99WO-US03307.
XX
XX      18-MAR-1998; 98US-0040961.
XX      06-AUG-1998; 98US-0130114.
XX
XX      (PHAR-) PHARMACOPEDIA INC.
XX
XX      Damaj BB, Horlick KA, Kobbins AK;
XX
XX      WPI; 1999-610610/52.
XX      P-PSDB; AAY28843.
XX
XX      New method for expressing genes from recombinant eukaryotic cells,
XX      useful for gene therapy
XX
XX      Claim 24; Fig 2; 86pp; English.
XX
XX      The present sequence is a DNA encoding Epstein Barr Virus Nuclear
XX      Antigen 1 (EBNA 1), which is obtained from commercially available
XX      plasmid pCMVEBNA. EBNA 1 protein is used to stably maintain episomes
XX      containing EBV origin of replication (orip) and a gene encoding
XX      protein or RNA of interest. Eucaryotic host cells expressing EBNA 1
XX      protein are transfected with these episomes to produce recombinant
XX      cell lines expressing multiple genes of interest. This provides a
XX      rapid and reliable method of stably expressing multiple genes in
XX      transfected cells. The episomes are useful in the transfection of genes
XX      encoding receptors, transporter proteins, ion channels, adhesion
XX      molecules and transcription factors. The episomes carrying desired genes
XX      can also be used to transfect cells in gene therapy, antisense therapy,
XX      for gene amplification, cell immortalisation, etc.
XX
XX      Sequence 1925 BP; 486 A; 352 C; 872 G; 215 T; 0 other;
SQ

```

```

Query Match      85.0%; Score 17; DB 20; Length 1925;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      4 acagacaatggactccc 20
        |||
Db      1679 ACAGACAATGGACTCCC 1663

```

```

RESULT 8
AAA50254/C
ID      AAA50254 standard; DNA; 1926 BP.
XX
XX      AAA50254;
XX
XX      07-NOV-2000 (first entry)
XX
XX      Epstein Barr virus nuclear antigen 1 protein (EBNA1) DNA.
XX
XX      EBV; nuclear antigen 1; EBNA1; episome; transfection; selection;
XX      gene therapy; ds.
XX
XX      Epstein-barr virus.
XX
XX      WO200047778-A1.
XX
XX      17-AUG-2000.
XX
XX      11-FEB-2000; 2000WO-US03547.
XX
XX      11-FEB-1999; 99US-0249585.
XX

```

XX (PHAR-) PHARMACOPEIA INC.  
 XX Horlick RA, Chelsky D;  
 XX WPI: 2000-515062/46.  
 XX P-PSDB; AAY95856.  
 XX Stably transfecting eukaryotic cells with at least one episome for the  
 XX production of a desired protein in vitro and for gene therapy.  
 XX Disclosure; Fig 2; 53pp; English.  
 XX  
 XX The present sequence is that of DNA encoding the Epstein-Barr virus  
 XX (EBV) nuclear antigen protein 1 (EBNA1, see AAY95856). EBNA1 is  
 XX utilised in a novel method for obtaining a eukaryotic cell that is  
 XX stably transfected with at least one episome. This method involves  
 XX transfecting a eukaryotic cell with: (1) a first episome comprising  
 XX an EBV origin of replication (oriP, see AAA50253), a gene encoding a  
 XX first protein whose expression results in cell death and a  
 XX selectable marker for eukaryotic cells; and (2) a second episome  
 XX comprising an EBV oriP and a gene encoding a second protein, where  
 XX expression of the second protein prohibits the occurrence of cell  
 XX death resulting from expression of the first protein to produce  
 XX doubly transfected cells which also express an antigen that  
 XX promotes retention of the episomes by the cells. The doubly  
 XX transfected cells are maintained under conditions in which the  
 XX first and second proteins and the selectable marker are expressed,  
 XX and the selective pressure specified by the marker is maintained.  
 XX Under these conditions, only cells containing both episomes live.  
 XX Preferably, EBNA1 is expressed from 1 of the episomes, and the  
 XX protein of interest from the other episome. Either or both  
 XX epitopes may further comprise a nucleic acid sequence encoding a  
 XX protein desired to be expressed in the cell (e.g. a therapeutic  
 XX protein), a nucleic acid encoding an RNA that is not intended to  
 XX be translated (e.g. a therapeutic RNA), or a DNA sequence used as  
 XX a tag for the cells. The method is applicable to cell culture or  
 XX intact organisms, for gene therapy. It allows the rapid  
 XX establishment of eukaryotic cells that stably and reliably express  
 XX a gene of interest, using a novel method of selection, and  
 XX maintenance of that selection without the need for exogenous  
 XX selection factors, such as antibiotics.  
 XX  
 XX Sequence 1926 BP; 487 A; 352 C; 872 G; 215 T; 0 other;  
 SQ

Query Match 85.0%; Score 17; DB 21; Length 1926;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 acagacaaatggactccc 20  
 |||||  
 Db 1680 ACAGACAATGGACTCCC 1664

RESULT 9  
 AAF82902/c  
 ID AAF82902 standard; DNA; 1926 BP.  
 XX  
 XX AAF82902;  
 XX  
 XX 29-JUN-2001 (first entry)  
 XX  
 XX EBV tethering protein EBNA1 encoding DNA.  
 XX  
 XX Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;  
 XX Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;  
 XX EBV; latency-associated nuclear antigen; LANA; EBNA1; ds.  
 XX  
 XX Epstein-barr virus.  
 XX  
 XX Key Location/Qualifiers  
 XX CDS 1..1926

FT /\*tag- a  
 XX WO200125484-A2.  
 XX 12-APR-2001.  
 XX 29-SEP-2000; 2000WO-US26908.  
 XX 01-OCT-1999; 99US-0410399.  
 XX (UNMI ) UNIV MICHIGAN.  
 XX Robertson ES, Colter MA;  
 XX WPI: 2001-281736/29.  
 XX P-PSDB; AAB62332.  
 XX A composition for use in gene therapy comprises an expression vector  
 XX PT that includes a nucleic acid sequence encoding a nucleic acid binding  
 XX protein.  
 XX Disclosure; Fig 9C; 60pp; English.  
 XX  
 XX The invention provides a composition comprising nucleic acid, histone H1  
 XX protein and expression vector operationally encoding a protein suitable  
 XX for tethering the nucleic acid to the histone H1 protein, where the  
 XX tethering protein is LANA. The composition is useful in aiding the  
 XX retention of the viral DNA in the host cell. The viral vector encodes a  
 XX protein suitable for tethering DNA to histone H1. Methods for screening  
 XX for compounds which are agonistic or antagonistic for the tethering of  
 XX viral proteins to histone H1 and DNA binding sites are useful for  
 XX developing the method of viral transfer. The composition has applications  
 XX to gene therapy, including the treatment of multiple sclerosis,  
 XX Parkinson's disease, Huntington disease and diabetes. The present  
 XX sequence represents the nucleotide sequence of the Epstein-barr virus  
 XX (EBV) tethering protein ENNA1.  
 XX  
 XX Sequence 1926 BP; 487 A; 352 C; 872 G; 215 T; 0 other;  
 SQ

Query Match 85.0%; Score 17; DB 22; Length 1926;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 acagacaaatggactccc 20  
 |||||  
 Db 1680 ACAGACAATGGACTCCC 1664

RESULT 10  
 AAX90923  
 ID AAX90923 standard; DNA; 5452 BP.  
 XX  
 XX AAX90923;  
 XX  
 XX 17-JAN-2000 (first entry)  
 XX  
 XX Anti-sense strand of pCMVEBNA plasmid.  
 XX  
 XX Anti-sense strand; plasmid pCMVEBNA; EBNA 1; episome; transfection;  
 XX Epstein Barr Virus Nuclear Antigen 1; origin of replication;  
 XX EBV oriP; eucaryotic host cell; recombinant cell line; ion channel;  
 XX multiple gene expression; receptor; transporter protein; gene therapy;  
 XX transcription factor; adhesion molecule; antisense therapy;  
 XX gene amplification; cell immortalisation; ss.  
 XX  
 XX Epstein-barr virus.  
 XX  
 XX Cytomegalovirus.  
 XX Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 XX CDS complement (3032..4957)  
 XX /\*tag- a

```

FT FT      /product= "EBNA 1 protein"
XX XX      /note= "Epstein Barr Virus Nuclear Antigen 1"
PN PN      WO9947647-A1.
PD PD      23-SEP-1999.
PF PF      12-FEB-1999; 99WO-US03307.
XX XX      18-MAR-1998; 98US-0040961.
PR PR      06-AUG-1998; 98US-0130114.
XX XX      (PHAR-) PHARMACOPEIA INC.
PI PI      DanaJ BB, Horlick RA, Robbins AK;
XX XX      WPI; 1999-610610/52.
DR DR      New method for expressing genes from recombinant eukaryotic cells,
XX XX      useful for gene therapy
PT PT      Example 1; Fig 1; 86pp; English.
PS PS
XX XX      The present sequence is an anti-sense strand of commercially available
CC CC      plasmid pCMVEBNA from which Epstein Barr Virus Nuclear Antigen 1 (EBNA
CC CC      1) encoding DNA is obtained. EBNA 1 protein is used to stably maintain
CC CC      episomes containing EBV origin of replication (oriP) and a gene encoding
CC CC      protein or RNA of interest. Eucaryotic host cells expressing EBNA 1
CC CC      protein are transfected with these episomes to produce recombinant
CC CC      cell lines expressing multiple genes of interest. This provides a
CC CC      rapid and reliable method of stably expressing multiple genes in
CC CC      transfected cells. The episomes are useful in the transfection of genes
CC CC      encoding receptors, transporter proteins, ion channels, adhesion
CC CC      molecules and transcription factors. The episomes carrying desired genes
CC CC      can also be used to transfect cells in gene therapy, antisense therapy,
CC CC      for gene amplification, cell immortalisation, etc.
XX XX      Sequence 5452 BP; 1108 A; 1736 C; 1246 G; 1362 T; 0 other;
SQ

```

Query Match 85.0%; Score 17; DB 20; Length 5452;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 4 acagacaatggactccc 20
Db 742 acagacaatggactccc 758

```

RESULT 11  
 ID AAQ40324/C  
 XX AAQ40324 standard; cDNA; 8575 BP.  
 XX AC AAQ40324;  
 XX DT 09-AUG-1993 (first entry)  
 XX DE Sequence of PHEB030.  
 XX KW Neurotrophic factor; growth promoting activity; GPA; ss.  
 XX OS Gallus domesticus.  
 XX FH Key Location/Qualifiers  
 FT FT misc\_feature 781..3061  
 FT FT /tag= a  
 FT FT repeat\_region 882..1474  
 FT FT /tag= b  
 FT FT misc\_feature 2587..2665  
 FT FT /tag= c  
 FT FT /label= dyad region  
 FT FT misc\_feature 3062

```

FT FT      /tag= d
FT FT      /note= "HSV TK Term 3' end"
FT FT      /tag= e
FT FT      /label= CMV
FT FT      3738..2901
FT FT      /tag= f
FT FT      2902..3988
FT FT      /tag= g
FT FT      /label= SP6
FT FT      3989..3993
FT FT      /tag= h
FT FT      /note= "SP6 RNA start"
FT FT      4510..4646
FT FT      /tag= i
FT FT      /label= SV40 early poly A
FT FT      /note= "g"
FT FT      4647..4691
FT FT      /tag= j
FT FT      /label= SV40 origin
FT FT      4692
FT FT      /tag= k
FT FT      /label= HSV TK Term 5' end
FT FT      complement (5185..6207)
FT FT      /tag= l
FT FT      /label= HPG gene from PLG89-- hygroresistance
FT FT      complement (6234..6279)
FT FT      /tag= m
FT FT      /label= HSV TK
FT FT      complement (6306..6359)
FT FT      /tag= n
FT FT      complement (6360..639)
FT FT      /tag= o
FT FT      complement (6480..6803)
FT FT      /tag= p
FT FT      /label= PBR 322; tet-Roepptide
FT FT      complement (7604..8448)
FT FT      /tag= q
FT FT      /label= beta lastanase
XX XX      WO9307270-A.
PN PN      15-APR-1993.
PD PD
XX XX      29-SEP-1992; 92WO-US08258.
PF PF
XX XX      01-OCT-1991; 91US-0769622.
PR PR      (GETH ) GENENTECH INC.
XX XX      (UYOR-) UNIV OREGON HEALTH SCI.
PA PA
XX XX      Cachianes G, Eckenstein FP, Leung D, Nishi R;
PI PI
XX XX      WPI; 1993-134459/16.
DR DR
XX XX      Nucleic acid encoding growth promoting activity - used for
PT PT      developing prods. for diagnosis and treatment of neurological
PT PT      diseases and disorders
XX XX      Disclosure; Fig 4; 74pp; English.
PS PS
XX XX      Eyes from embryonic day 15 (E15) chicken embryos were dissected and
CC CC      total RNA was isolated. The RNA was used to prepare cDNA which was
CC CC      then ligated to SfiI digested lambda-HEB0 vector. The E15 cDNA
CC CC      library was screened using oligonucleotide probes based on the
CC CC      partial amino acid sequences of GPA obtd. by microsequencing of 3
CC CC      different peptide fragments of purified GPA. The oligonucleotides
CC CC      are referred to as o-GPA-1, o-GPA-2 and o-GPA-3. Lambda CE15 #19
CC CC      hybridized with all three probes. The expression of GPA involves
CC CC      the use of a plasmid expression vector (PHEB030) containing the oriP
CC CC      region from EBV. PHEB030 comprises the strong CMV promoter, a multiple
CC CC      cloning region for insertion of foreign (exogenous) genes downstream
CC CC      of the CMV promoter, the oriP region of EBV for plasmid replication

```

CC in host cells expressing EBNA-1, a hygromycin resistance gene for  
 CC selection in eukaryotes, the origin of replication from pBR322 for  
 CC replication in prokaryotes, and an ampicillin resistance gene for  
 CC selection in prokaryotes. In the sequence "N" is used to  
 CC designate the nucleotides that comprise the arbitrary 350 base pair  
 CC cDNA insert in pHEB030.

XX Sequence 8575 BP; 1989 A; 2167 C; 2032 G; 2037 T; 350 other;

Query Match 85.0%; Score 17; DB 14; Length 8575;

Best Local Similarity 100.0%; Pred. No. 38;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 acagacaatgactccc 20

|||||

Db 234 ACAGACATGGACTCCC 218

RESULT 12

AAZ23778  
 ID AAZ23778 standard; DNA; 8705 BP.

XX

AC AAZ23778;

DT 14-JAN-2000 (first entry)

XX

DE Vector pShuttle DNA.

XX Antisense; DNA library; identification; multiple cloning site; MCS;

KW inhibition; ss.

XX Synthetic.

OS WO9950457-A1.

XX

PN 07-OCT-1999.

XX

XX 28-MAR-1999; 99WO-US06742.

XX

PR 28-MAR-1998; 98US-0079792.

XX

PR 06-NOV-1998; 98US-0107504.

XX

PA (UTAH ) UNIV UTAH RES FOUND.

XX

XX Ruffner DE, Pierce ML, Chen Z;

XX

XX WPI; 1999-610866/52.

XX

DR Production of antisense libraries, used for identifying antisense

XX agents and for identifying target sites for antisense-mediated

XX inhibition of a selected gene -

XX Claim 16; Page 43-50; 63pp; English.

XX

XX This invention describes a novel method for generating an antisense

XX library targeted to a selected RNA transcript. The methods can be used

XX for identifying antisense agents and for identifying target sites for

XX antisense-mediated inhibition of a selected gene. The use of a direct

XX library for target site selection significantly simplifies the screening

XX process, since only very small libraries need be prepared and assayed.

XX This sequence represents the vector pShuttle which is used in the method

XX of the invention.

XX

XX Sequence 8705 BP; 1875 A; 2575 C; 2078 G; 2177 T; 0 other;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Db 6607 acagacaatgactccc 6623

RESULT 13

AAV21683/c

ID AAV21683 standard; DNA; 9600 BP.

XX

AC AAV21683;

XX

DT 17-AUG-1998 (first entry)

XX

DE Vector plasmid pCMVkmITR-EPI.

XX

XX Polynucleotide delivery; plasmid pCMVkmITR-EPI; vector;

XX gene therapy; vaccine; polycationic agent; ss.

XX

OS Chimeric - Epstein-Barr virus.

OS Chimeric - Adeno-associated virus.

OS Chimeric - Cytomegalovirus.

OS Chimeric - Bos taurus.

XX

XX Key Location/Qualifiers

FT CDS 14..2594

FT /\*tag= a

FT /product= "EBV nuclear antigen A"

FT 2623..4559

FT /\*tag= b

FT /note= "EBV origin of replication"

FT 4928..5104

FT repeat\_unit

FT /\*tag= c

FT /rpt\_type= INVERTED

FT /note= "AAV inverted terminal repeat"

FT 7189..7355

FT repeat\_unit

FT /\*tag= d

FT /rpt\_type= INVERTED

FT /note= "AAV inverted terminal repeat"

FT 5112..6734

FT promoter

FT /\*tag= e

FT /note= "CMV immediate-early enhancer/promoter"

FT 6818..7050

FT terminator

FT /\*tag= f

FT /note= "bovine growth hormone polyA sequence"

XX

XX WO9806437-A2.

XX

XX 19-FEB-1998.

XX

XX 13-AUG-1997; 97WO-US14465.

XX

XX 13-AUG-1996; 96US-0023867.

XX

XX (CHIR ) CHIRON CORP.

XX

XX Cohen F, Dubois-Stringfellow N, Dwarki V, Innis MA;

XX Murphy JE, Tetsuo U, Zukermann R;

XX

XX WPI; 1998-159296/14.

XX

XX Polycationic agents based on alpha-amino acids, able to complex

XX with nucleic acid - to facilitate its entry - to cell, condense it

XX and protect it against serum degradation, particularly for use in

XX gene therapy

XX

XX Disclosure; Page 77-80; 100pp; English.

XX

XX This polynucleotide comprises the DNA sequence of vector plasmid

XX pCMVkmITR-EPI, which contains an Epstein-Barr virus (EBV) origin

XX of replication from plasmid pCEP4, a coding region for EBV nuclear

XX antigen A from pCEP4, a pair of inverted terminal repeats from

XX adeno-associated virus, a cytomegalovirus enhancer/promoter, a

XX bovine growth hormone polyA sequence, and a kanamycin resistance

XX selectable marker. Polynucleotides encoding polypeptides, such as

XX erythropoietin or leptin, and ribozymes and antisense

XX

XX

XX

XX

XX

CC polynucleotides can be inserted into the vector. The vector is  
 CC preferred for use in novel compositions and methods for improved  
 CC polynucleotide delivery into cells. In these methods, polycationic  
 CC agents are used to increase the frequency of uptake of a  
 CC nucleic acid (see also AAV21684-86) into a cell. The polycationic  
 CC agent can condense with the nucleic acid and inhibit serum and/or  
 CC nuclease degradation of the nucleic acid. The nucleic acid can be  
 CC a vector, may express a therapeutic protein or a vaccinating viral  
 CC or cancer antigen, or is itself therapeutic (antisense or  
 CC ribozyme). The methods and compositions can be used in the gene  
 CC therapy of many diseases.  
 XX  
 SQ Sequence 9600 BP; 2326 A; 2376 C; 2817 G; 2081 T; 0 other;

Query Match 85.0%; Score 17; DB 19; Length 9600;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 acagacaaatggactccc 20  
 |||||  
 DB 2109 ACAGACAATGGACTCCC 2093

RESULT 14  
 AAA53873/c  
 ID AAA53873 standard; DNA; 9725 BP.  
 XX  
 AC AAA53873;

XX 03-JAN-2001 (first entry)  
 DT  
 DE Expression vector pRIG19.

XX Vector; endogenous gene; activation; over-expression;  
 KW erythropoietin; growth hormone; drug discovery;  
 KW granulocyte colony stimulating factor; ds.

XX Synthetic.

XX WO2000049162-A2.

XX 24-AUG-2000.

XX 22-FEB-2000; 2000WO-0504429.

XX 19-FEB-1999; 99US-0253022.

XX 08-MAR-1999; 99US-0263814.

XX 26-MAR-1999; 99US-0276820.

XX (ATHE-) ATHERSYS INC.

XX Harrington JJ, Sherf B, Rundlett S;

XX WPI; 2000-549276/50.

XX Non-targeted activation of endogenous genes, e.g. for the production of  
 PT erythropoietin, growth hormone or granulocyte-colony stimulating factor  
 PT proteins and for drug discovery

XX Disclosure; Fig 30; 240pp; English.

XX New methods, vectors and cells are described for non-targeted  
 CC activation and over-expression of endogenous genes. The vector  
 CC constructs comprise transcriptional regulatory sequences (TRS) and  
 CC unpaired splice donor sequences (USDS), preferably the vectors  
 CC comprise (in sequential order) a TRS, an USDS, a rare cutting  
 CC restriction site (RCS) and a linearization site (LS) with a second  
 CC TRS linked to a selectable marker (SM) lacking a polyadenylation  
 CC signal. The methods, vectors and cells comprising the vectors may  
 CC be used for the non-targeted activation and over-expression of  
 CC endogenous genes, e.g. for the production of proteins (including  
 CC erythropoietin, growth hormone or granulocyte-colony stimulating

CC factor) and drug discovery. The advantage of these methods are that  
 CC endogenous genes including those associated with human disease and  
 CC development, may be activated and isolated without prior knowledge  
 CC of the sequence structure, function or expression profile of  
 CC the genes being known.

SQ Sequence 9725 BP; 2425 A; 2507 C; 2448 G; 2343 T; 2 other;

Query Match 85.0%; Score 17; DB 21; Length 9725;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 acagacaaatggactccc 20  
 |||||  
 DB 3331 ACAGACAATGGACTCCC 3315

RESULT 15  
 AAA53879/c  
 ID AAA53879 standard; DNA; 9732 BP.  
 XX  
 AC AAA53879;

XX 03-JAN-2001 (first entry)

XX Expression vector pRIG-TP.

XX Vector; endogenous gene; activation; over-expression;  
 KW erythropoietin; growth hormone; drug discovery;  
 KW granulocyte colony stimulating factor; ds.

XX Synthetic.

XX WO2000049162-A2.

XX 24-AUG-2000.

XX 22-FEB-2000; 2000WO-0504429.

XX 19-FEB-1999; 99US-0253022.

XX 08-MAR-1999; 99US-0263814.

XX 26-MAR-1999; 99US-0276820.

XX (ATHE-) ATHERSYS INC.

XX Harrington JJ, Sherf B, Rundlett S;

XX WPI; 2000-549276/50.

XX Non-targeted activation of endogenous genes, e.g. for the production of  
 PT erythropoietin, growth hormone or granulocyte-colony stimulating factor  
 PT proteins and for drug discovery

XX Example 15; Fig 37; 240pp; English.

XX New methods, vectors and cells are described for non-targeted  
 CC activation and over-expression of endogenous genes. The vector  
 CC constructs comprise transcriptional regulatory sequences (TRS) and  
 CC unpaired splice donor sequences (USDS), preferably the vectors  
 CC comprise (in sequential order) a TRS, an USDS, a rare cutting  
 CC restriction site (RCS) and a linearization site (LS) with a second  
 CC TRS linked to a selectable marker (SM) lacking a polyadenylation  
 CC signal. The methods, vectors and cells comprising the vectors may  
 CC be used for the non-targeted activation and over-expression of  
 CC endogenous genes, e.g. for the production of proteins (including  
 CC erythropoietin, growth hormone or granulocyte-colony stimulating  
 CC factor) and drug discovery. The advantage of these methods are that  
 CC endogenous genes including those associated with human disease and  
 CC development, may be activated and isolated without prior knowledge  
 CC of the sequence structure, function or expression profile of  
 CC the genes being known.

XX

SQ Sequence 9732 BP; 2427 A; 2512 C; 2443 G; 2348 T; 2 other;

Query Match 85.0%; Score 17; DB 21; Length 9732;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 aragacaatggactccc 20  
 |||  
 Db 3338 ACAGACAATGGACTCCC 3322

Search completed: May 18, 2002, 17:45:53  
 Job time: 13742 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2002, 14:58:03 ; Search time 2878.96 Seconds  
(without alignments)  
145.376 Million cell updates/sec

Title: US-09-623-329-4

Perfect score: 20

Sequence: 1 aatacagacaatggactccc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_on.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_on.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vl.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
------------	-------	-------	-------	--------	----	----	-------------

1	20	100.0	20	6	AX018445
2	17.4	87.0	185076	2	AC102989
3	17.4	87.0	219489	2	AC105615
4	17	85.0	26	14	AX045700
5	17	85.0	627	14	HHU21193
6	17	85.0	627	14	HHU21194
7	17	85.0	627	14	HHU21195
8	17	85.0	627	14	HHU21196
9	17	85.0	627	14	HHU21197
10	17	85.0	627	14	HHU21198
11	17	85.0	627	14	HHU21199
12	17	85.0	627	14	HHU21200
13	17	85.0	627	14	HHU21201
14	17	85.0	627	14	HHU21202
15	17	85.0	627	14	HHU21203
16	17	85.0	627	14	HHU21204
17	17	85.0	636	14	HHU21205
18	17	85.0	1748	6	AR116267
19	17	85.0	1926	6	AX107940
20	17	85.0	5452	6	AR083151
21	17	85.0	5452	12	U02454
22	17	85.0	9600	6	A92665
23	17	85.0	9600	6	ARI58345
24	17	85.0	10596	6	I25041
25	17	85.0	10596	6	I30503
26	17	85.0	10737	12	XXU02428
27	17	85.0	10850	12	U02455
28	17	85.0	11265	6	AX031159
29	17	85.0	172281	14	EBV
30	17	85.0	184113	14	HS4B958RAJ
31	16.8	84.0	397	6	AX071548
32	16.8	84.0	2099	10	AF304376
33	16.8	84.0	10877	6	E30086
34	16.8	84.0	10877	9	AF210054
35	16.8	84.0	39124	9	AC004186
36	16.8	84.0	62456	9	AL662782
37	16.8	84.0	87589	2	AC066602
38	16.8	84.0	87589	2	AC026211
39	16.8	84.0	91801	2	AC095906
40	16.8	84.0	95097	2	AC106959
41	16.8	84.0	100000	9	AP000517
42	16.8	84.0	101473	9	HSJ813D12
43	16.8	84.0	107137	9	AC011292
44	16.8	84.0	125963	2	AL671855
45	16.8	84.0	137111	9	AC007690

#### ALIGNMENTS

RESULT	1	AX018445	Sequence 4 from Patent WO9945155.	20 bp	DNA	Linear	PAT 07-SEP-2000
LOCUS	AX018445	Sequence 4 from Patent WO9945155.					
DEFINITION	AX018445	Sequence 4 from Patent WO9945155.					
ACCESSION	AX018445	Sequence 4 from Patent WO9945155.					
VERSION	AX018445.1	GI:10042596					
KEYWORDS	Epstein-Barr virus.						
SOURCE	Human herpesvirus 4						
ORGANISM	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;						
	Gammaherpesvirinae; Lymphocryptovirus.						
REFERENCE	1 (bases 1 to 20).						
AUTHORS	Middelcorp,J.M., Van Den Brule,A.J. and Vervoot,M.B.						
TITLE	Oligonucleotides for the amplification and detection of Epstein						
	barr virus (ebv) nucleic acid						
JOURNAL	Patent: WO 945155-A 4 10-SEP-1999;						
	MIDDELDRP JAAP MICHEL (NL); AKZO NOBEL NV (NL); DEN BRULE						
	ADRIANUS JOHANNES CH (NL); VERVOORT MARCEL BARTOLINA HEND (NL)						
FEATURES	Location/Qualifiers						
source	1..20						
	/organism="Human herpesvirus 4"						
	/db_xref="taxon:10376"						
BASE COUNT	8 a	6 c	3 g	3 t			

ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 20;

Best Local Similarity 100.0%; Pred. No. 1.6;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 aatacaacaatgactccc 20

Db 1 AATACAGACAATGACTCCC 20

RESULT 2

AC102989

LOCUS Rattus norvegicus clone CH230-57J23, linear HTG 20-DEC-2001

DEFINITION \*\*\* 98 unordered pieces.

AC102989

VERSION HTG; HTGS\_PHASE1

KEYWORDS Norway rat.

SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelesostomi;

ORGANISM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 185076)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,

Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbara,J.,

Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,

Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,

Cartier,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,

Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,

Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,

Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,

Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,

Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,

Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,

Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,

Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S.,

Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,

Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,

Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,

Louisege,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,

Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,

Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,

Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,

Morgan,M., Morris,M., Moser,M., Neal,D., Newton,J., Newton,N.,

Nguyen,M., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S.,

Ogum,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,

Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,

Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,

Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,

Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,

Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,

Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,

Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,

Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,

Wallington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,

Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,

Weinstock,G. and Gbbs,R.

Direct Submission

Unpublished

2 (bases 1 to 185076)

Worley,K.C.

Direct Submission

Submitted (26-NOV-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Dec 20, 2001 this sequence version replaced gi:17066855.

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

Project Information

Center project name: GHEI

Center clone name: CH230-57J23

Summary Statistics

Assembly program: Phrap; version 0.990329First call to

findPhrapList

Consensus quality: 161943 bases at least Q40

Consensus quality: 175270 bases at least Q30

Consensus quality: 184561 bases at least Q20

Estimated insert size: 141688; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-gel estimation

Quality coverage: 1.8x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

NOTE: This is a 'working draft' sequence. It currently

consists of 98 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 3745: contig of 3745 bp in length

3746: gap of unknown length

8327: contig of 4482 bp in length

8428: contig of 3903 bp in length

12430: gap of unknown length

14487: contig of 2057 bp in length

14587: gap of unknown length

18263: contig of 3676 bp in length

18363: gap of unknown length

21906: contig of 3543 bp in length

22006: gap of unknown length

25145: contig of 3139 bp in length

25245: gap of unknown length

28446: contig of 3201 bp in length

28447: gap of unknown length

31271: contig of 2725 bp in length

31371: gap of unknown length

33241: contig of 1870 bp in length

33341: gap of unknown length

35692: contig of 2351 bp in length

35792: gap of unknown length

38776: contig of 2984 bp in length

38876: gap of unknown length

41421: contig of 2545 bp in length

41521: gap of unknown length

44738: contig of 3207 bp in length

44828: gap of unknown length

46476: contig of 1648 bp in length

46576: gap of unknown length

48937: contig of 2361 bp in length

49037: gap of unknown length

51773: contig of 2736 bp in length

51873: gap of unknown length

53593: contig of 1720 bp in length

53693: gap of unknown length

55913: contig of 2220 bp in length

56013: gap of unknown length

57776: contig of 1763 bp in length

57876: gap of unknown length

59339: contig of 1463 bp in length

59439: gap of unknown length

61076: contig of 1637 bp in length

61176: gap of unknown length



```

* 61177 63414: contig of 2238 bp in length
* 63415 63514: gap of unknown length
* 63515 65060: contig of 1546 bp in length
* 65061 65160: gap of unknown length
* 65161 67391: contig of 2231 bp in length
* 67392 67491: gap of unknown length
* 67492 69718: contig of 2227 bp in length
* 69719 69818: gap of unknown length
* 69819 72411: contig of 2593 bp in length
* 72412 72511: gap of unknown length
* 72512 74417: contig of 1906 bp in length
* 74418 74517: gap of unknown length
* 74518 76054: contig of 1537 bp in length
* 76055 76154: gap of unknown length
* 76155 78148: contig of 1994 bp in length
* 78149 78248: gap of unknown length
* 78249 79854: contig of 1606 bp in length
* 79855 79954: gap of unknown length
* 79955 81481: contig of 1527 bp in length
* 81482 81581: gap of unknown length
* 81582 83146: contig of 1565 bp in length
* 83147 83246: gap of unknown length
* 83247 85203: contig of 1957 bp in length
* 85204 85303: gap of unknown length
* 85304 86880: contig of 1577 bp in length
* 86881 86980: gap of unknown length
* 86981 89616: contig of 2636 bp in length
* 89617 89716: gap of unknown length
* 89717 91551: contig of 1835 bp in length
* 91552 91651: gap of unknown length
* 91652 93129: contig of 1478 bp in length
* 93130 93229: gap of unknown length
* 93230 94356: contig of 1127 bp in length
* 94357 95984: gap of unknown length
* 95985 96084: contig of 1528 bp in length
* 96085 97420: contig of 1336 bp in length
* 97421 97520: gap of unknown length
* 97521 99139: contig of 1619 bp in length
* 99140 99239: gap of unknown length
* 99240 100669: contig of 1430 bp in length
* 100670 102501: gap of unknown length
* 102502 102601: contig of 1732 bp in length
* 102602 103963: gap of unknown length
* 103964 104063: contig of 1362 bp in length
* 104064 106409: gap of unknown length
* 106410 106509: contig of 2346 bp in length
* 106510 107768: gap of unknown length
* 107769 107868: contig of 1259 bp in length
* 107869 109017: gap of unknown length
* 109018 109117: contig of 1149 bp in length
* 109118 110893: gap of unknown length
* 110894 112466: contig of 1776 bp in length
* 112467 112566: gap of unknown length
* 112567 114636: contig of 2070 bp in length
* 114637 114736: gap of unknown length
* 114737 116011: contig of 1275 bp in length
* 116012 116111: gap of unknown length
* 116112 117673: contig of 1562 bp in length
* 117674 117773: gap of unknown length
* 117774 119486: contig of 1713 bp in length
* 119487 119586: gap of unknown length
* 119587 121621: contig of 2035 bp in length
* 121622 121721: gap of unknown length
* 121722 124577: contig of 2856 bp in length
* 124578 124677: gap of unknown length

```

Query Match 87.0%; Score 17.4; DB 2: Length 185076;  
 Best Local Similarity 94.7%; Pred. No. 54;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 aatacagacaatgactcc 19  
 Db 88493 AATCAGACAAATGACTCC 88511

## RESULT 3

AC105615 219489 bp DNA linear HTG 09-JAN-2002  
 LOCUS Rattus norvegicus chromosome Chr18 clone CH230-234J12, WORKING  
 DEFINITION DRAFT SEQUENCE, 33 unordered pieces.

AC105615 1 GI:18092838  
 VERSION HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS Norway rat.  
 SOURCE

## ORGANISM

Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

## REFERENCE:

## AUTHORS

1 (bases 1 to 219489)  
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,  
 Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Banks, T., Barbaria, J.,  
 Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J.,  
 Bowie, S., Briefva, M., Brown, E., Brown, M., Bryant, N.P., Buha, C.,  
 Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,  
 Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,  
 Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,  
 Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,  
 Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,  
 Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,  
 Dugan-Rochia, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,  
 Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,  
 Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,  
 Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,  
 Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,  
 Hernandez, J., Hernaandez, O., Hodgson, A., Hoques, M., Holloway, C.,  
 Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,  
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S.,  
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,  
 Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,  
 Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,  
 Loulsegod, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,  
 Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M.,  
 Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Monabbat, K.,  
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,  
 Oquh, M., Okwuonu, G., Otagunye, N., Oviedo, R., Pace, A., Payton, B.,  
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,  
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,  
 Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoostari, N.,  
 Stinson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H.,  
 Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,  
 Tang, H., Tauney, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,  
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,  
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
 Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S.,  
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G. and Gibbs, R.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## JOURNAL

## JOURNAL

## JOURNAL

## JOURNAL

## JOURNAL

## JOURNAL

## JOURNAL

## JOURNAL

## JOURNAL

## JOURNAL

## JOURNAL

## JOURNAL

## JOURNAL

## JOURNAL

## JOURNAL

## JOURNAL

Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 ----- Project Information  
 ----- Project name: GNIQ

```

Center clone name: CH230-234J12
----- Summary Statistics -----
Assembly Program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 198171 bases at least Q40
Consensus quality: 202505 bases at least Q30
Consensus quality: 205741 bases at least Q20
Estimated insert size: 207158; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 3.5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

Query Match 87.0%; Score 17.4; DB 2; Length 219489;  
Best Local Similarity 94.7%; Pred. NO. 54;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 atacagacaatggactccc 20  
|||||  
Db 6014 ATACAGACAAGGACTCCC 6032

RESULT 4  
LOCUS AX045700 V6 bp DNA linear PAT 24-NOV-2000  
DEFINITION Sequence 14 from Patent WO0066769.  
ACCESSION AX045700  
VERSION AX045700.1 GI:11344070  
KEYWORDS synthetic construct.  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1. (bases 1 to 26)  
AUTHORS Ng,R.H., Daykin,V. and Phillips,J.  
TITLE System and method for screening of nasopharyngeal carcinoma  
JOURNAL Patent: WO 0066769-A 14 09-NOV-2000;  
Advance Sentry Corporation (CA)

FEATURES  
Source Location/Qualifiers  
1..26  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/mole="oligonucleotide" 5' t  
BASE COUNT 10 a 7 c 4 g  
ORIGIN

Query Match 85.0%; Score 17; DB 6; Length 26;  
Best Local Similarity 100.0%; Pred. No. 86;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 acagacaatggactccc 20  
|||||  
Db 5 ACAGACAATGGACTCCC 21

RESULT 5  
LOCUS HHU21193/c  
DEFINITION Human herpesvirus 4, isolate AM nuclear antigen EBNA-1 gene,  
partial cds, C-terminal unique domain.  
ACCESSION U21193  
VERSION U21193.1 GI:710371  
KEYWORDS

```

SOURCE
ORGANISM      Epstein-Barr virus.
REFERENCE     Viruses; dsDNA viruses, no RNA stage: Herpesviridae;
AUTHORS      Gammaherpesvirinae; Lymphocryptovirus.
TITLE        1 (bases 1 to 609)
              Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,
              Rooney,C.M. and Arrand,J.R.
              Antigenic and sequence variation in the C-terminal unique domain of
              the Epstein-Barr virus nuclear antigen EBNA-1
JOURNAL      Virology (1995) In press
REFERENCE     2 (bases 1 to 627)
              Pepper,S.D.
              Direct Submission
              Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for
              Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,
              Manchester, Britain, M20 9BX
FEATURES     Location/Qualifiers
              source
                1..627
                /organism="Human herpesvirus 4"
                /isolate="AM"
                /specific_host="Homo sapiens"
                /db_xref="taxon:10376"
                <1..612
                /note="C-terminal unique domain"
                /codon_start=1
                /product="nuclear antigen EBNA-1"
                /protein_id="AA67275.1"
                /db_xref="GI:710372"
                /translation="ADDPGPGSTGPRGQDGGRRKKGWFGKIRGGGSKPKFENIA
                EGRLLARSHVETTDGTVAGVFYVGGSKTSLYNLRGIALAIPOCRLTPLSLRP
                EGMAGPGPGPLRESIVCYFIVFLQTHIFAEGLKDAIKDLVMPKPAPTCNIRVTV
                SFDDGVLDLPWFPMVFGAAAGDDGDDGGDEGDEGEQGE"
BASE COUNT   154 a 133 c 193 g 147 t
ORIGIN
Query Match 85.0%; Score 17; DB 14; length 627;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 acagacaatgactccc 20
    |||||
Db 366 ACAGACAATGGACTCCC 350

RESULT 7
HHU21195/c
LOCUS
DEFINITION   Human herpesvirus 4, isolate Ag876 nuclear antigen EBNA-1 gene,
ACCESSION   U21194
VERSION      U21194.1 GI:710373
KEYWORDS
SOURCE
ORGANISM     Human herpesvirus 4
Virus; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE    1 (bases 1 to 609)
              Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,
              Rooney,C.M. and Arrand,J.R.
              Antigenic and sequence variation in the C-terminal unique domain of
              the Epstein-Barr virus nuclear antigen EBNA-1
AUTHORS      the Epstein-Barr virus nuclear antigen EBNA-1
              Virology (1995) In press
TITLE        2 (bases 1 to 627)
              Pepper,S.D.
              Direct Submission
              Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for
              Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,
              Manchester, Britain, M20 9BX
FEATURES     Location/Qualifiers
              source
                1..627
                /organism="Human herpesvirus 4"
                /isolate="Ag876"
                /specific_host="Homo sapiens"
                /db_xref="taxon:10376"
                <1..612
                /note="C-terminal unique domain"
                /codon_start=1
                /product="nuclear antigen EBNA-1"
                /protein_id="AA67275.1"
                /db_xref="GI:710372"
                /translation="ADDPGPGSTGPRGQDGGRRKKGWFGKIRGGGSKPKFENIA
                EGRLLARSHVETTDGTVAGVFYVGGSKTSLYNLRGIALAIPOCRLTPLSLRP
                EGMAGPGPGPLRESIVCYFIVFLQTHIFAEGLKDAIKDLVMPKPAPTCNIRVTV
                SFDDGVLDLPWFPMVFGAAAGDDGDDGGDEGDEGEQGE"
BASE COUNT   154 a 133 c 193 g 147 t
ORIGIN
Query Match 85.0%; Score 17; DB 14; length 627;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 acagacaatgactccc 20
    |||||
Db 366 ACAGACAATGGACTCCC 350

RESULT 6
HHU21194/c
LOCUS
DEFINITION   Human herpesvirus 4, isolate Ag876 nuclear antigen EBNA-1 gene,
ACCESSION   U21194
VERSION      U21194.1 GI:710373
KEYWORDS
SOURCE
ORGANISM     Epstein-Barr virus.
Virus; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE    1 (bases 1 to 609)
              Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,
              Rooney,C.M. and Arrand,J.R.
              Antigenic and sequence variation in the C-terminal unique domain of
              the Epstein-Barr virus nuclear antigen EBNA-1
AUTHORS      the Epstein-Barr virus nuclear antigen EBNA-1
              Virology (1995) In press
TITLE        2 (bases 1 to 627)
              Pepper,S.D.
              Direct Submission
              Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for
              Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,
              Manchester, Britain, M20 9BX
FEATURES     Location/Qualifiers
              source
                1..627
                /organism="Human herpesvirus 4"
                /isolate="Ag876"
                /specific_host="Homo sapiens"
                /db_xref="taxon:10376"
                <1..612
                /note="C-terminal unique domain"
                /codon_start=1
                /product="nuclear antigen EBNA-1"
                /protein_id="AA67275.1"
                /db_xref="GI:710372"
                /translation="ADDPGPGSTGPRGQDGGRRKKGWFGKIRGGGSKPKFENIA
                EGRLLARSHVETTDGTVAGVFYVGGSKTSLYNLRGIALAIPOCRLTPLSLRP
                EGMAGPGPGPLRESIVCYFIVFLQTHIFAEGLKDAIKDLVMPKPAPTCNIRVTV
                SFDDGVLDLPWFPMVFGAAAGDDGDDGGDEGDEGEQGE"
BASE COUNT   154 a 132 c 195 g 146 t
ORIGIN
Query Match 85.0%; Score 17; DB 14; length 627;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 acagacaatgactccc 20
    |||||
Db 366 ACAGACAATGGACTCCC 350

RESULT 7
HHU21195/c
LOCUS
DEFINITION   Human herpesvirus 4, isolate EG nuclear antigen EBNA-1 gene,
ACCESSION   U21195
VERSION      U21195.1 GI:710375
KEYWORDS
SOURCE
ORGANISM     Epstein-Barr virus.
Virus; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE    1 (bases 1 to 609)
              Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,
              Rooney,C.M. and Arrand,J.R.
              Antigenic and sequence variation in the C-terminal unique domain of
              the Epstein-Barr virus nuclear antigen EBNA-1
AUTHORS      the Epstein-Barr virus nuclear antigen EBNA-1
              Virology (1995) In press
TITLE        2 (bases 1 to 627)
              Pepper,S.D.
              Direct Submission
              Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for
              Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,
              Manchester, Britain, M20 9BX
FEATURES     Location/Qualifiers
              source
                1..627
                /organism="Human herpesvirus 4"
                /isolate="EG"
                /specific_host="Homo sapiens"
                /db_xref="taxon:10376"
                <1..612
                /note="C-terminal unique domain"
                /codon_start=1
                /product="nuclear antigen EBNA-1"
                /protein_id="AA67277.1"
                /db_xref="GI:710376"
                /translation="ADDPGPGSTGPRGQDGGRRKKGWFGKIRGGGSKPKFENIA
                EGRLLARSHVETTDGTVAGVFYVGGSKTSLYNLRGIALAIPOCRLTPLSLRP
                EGMAGPGPGPLRESIVCYFIVFLQTHIFAEGLKDAIKDLVMPKPAPTCNIRVTV
                SFDDGVLDLPWFPMVFGAAAGDDGDDGGDEGDEGEQGE"
BASE COUNT   152 a 128 c 196 g 151 t
ORIGIN
Query Match 85.0%; Score 17; DB 14; length 627;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 acagacaatgactccc 20
    |||||
Db 366 ACAGACAATGGACTCCC 350

RESULT 7
HHU21195/c
LOCUS
DEFINITION   Human herpesvirus 4, isolate EG nuclear antigen EBNA-1 gene,
ACCESSION   U21195
VERSION      U21195.1 GI:710375
KEYWORDS
SOURCE
ORGANISM     Epstein-Barr virus.
Virus; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE    1 (bases 1 to 609)
              Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,
              Rooney,C.M. and Arrand,J.R.
              Antigenic and sequence variation in the C-terminal unique domain of
              the Epstein-Barr virus nuclear antigen EBNA-1
AUTHORS      the Epstein-Barr virus nuclear antigen EBNA-1
              Virology (1995) In press
TITLE        2 (bases 1 to 627)
              Pepper,S.D.
              Direct Submission
              Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for
              Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,
              Manchester, Britain, M20 9BX
FEATURES     Location/Qualifiers
              source
                1..627
                /organism="Human herpesvirus 4"
                /isolate="EG"
                /specific_host="Homo sapiens"
                /db_xref="taxon:10376"
                <1..612
                /note="C-terminal unique domain"
                /codon_start=1
                /product="nuclear antigen EBNA-1"
                /protein_id="AA67277.1"
                /db_xref="GI:710376"
                /translation="ADDPGPGSTGPRGQDGGRRKKGWFGKIRGGGSKPKFENIA
                EGRLLARSHVETTDGTVAGVFYVGGSKTSLYNLRGIALAIPOCRLTPLSLRP
                EGMAGPGPGPLRESIVCYFIVFLQTHIFAEGLKDAIKDLVMPKPAPTCNIRVTV
                SFDDGVLDLPWFPMVFGAAAGDDGDDGGDEGDEGEQGE"
BASE COUNT   154 a 132 c 195 g 146 t
ORIGIN
Query Match 85.0%; Score 17; DB 14; length 627;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 acagacaatgactccc 20
    |||||
Db 366 ACAGACAATGGACTCCC 350

```

```

Oy 4 acagacaatggactccc 20
    |||||||
Db 366 ACAGACAATGGACTCCC 350

RESULT 8
HHU21196/c
LOCUS
DEFINITION Human herpesvirus 4, isolate JF1 nuclear antigen EBNA-1 gene,
partial cds, C-terminal unique domain.
ACCESSION U21196
VERSION U21196.1 GI:710377
KEYWORDS
SOURCE
ORGANISM Epstein-Barr virus.
Human herpesvirus 4
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE 1 (bases 1 to 609)
AUTHORS Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,
Rooney,C.M. and Arrand,J.R.
TITLE Antigenic and sequence variation in the C-terminal unique domain of
the Epstein-Barr virus nuclear antigen EBNA-1
JOURNAL Virology (1995) In press
REFERENCE 2 (bases 1 to 627)
AUTHORS Pepper,S.D.
JOURNAL Direct Submission
Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for
Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,
Manchester, Britain, M20 9BX
FEATURES
    source
        Location/Qualifiers
            1..627
                /organism="Human herpesvirus 4"
                /isolate="JF1"
                /specific_host="Homo sapiens"
                /db_xref="taxon:10376"
                <1..612
                    /note="C-terminal unique domain"
                    /codon_start=1
                    /product="nuclear antigen EBNA-1"
                    /protein_id="AAA67278.1"
                    /db_xref="GI:710378"
                    /translation="ADDPGEGPSTGPRGGDGGRRKKGWFGKRRGGGSGNOKFENIA
                    EGRLLARCHEVETTEDGNVAGVFYVGGSKTSLYNLRGIGLAIPOCKLTPLSRIP
                    FGMAPGPGOPPLRESIVCYFVFLQTHIFAELKDAIKDLVLPKPAPTCNIRVTV
                    SFDDGVLDLPMPFPMVEGAAEGDGDGDEGDEGEQGE"
BASE COUNT 151 a 128 c 197 g 151 t
ORIGIN

Query Match 85.0%; Score 17; DB 14; Length 627;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CDS
Oy 4 acagacaatggactccc 20
    |||||||
Db 366 ACAGACAATGGACTCCC 350

RESULT 9
HHU21197/c
LOCUS
DEFINITION Human herpesvirus 4, isolate JF2 nuclear antigen EBNA-1 gene,
partial cds, C-terminal unique domain.
ACCESSION U21197
VERSION U21197.1 GI:710379
KEYWORDS
SOURCE
ORGANISM Epstein-Barr virus.
Human herpesvirus 4
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE 1 (bases 1 to 609)
AUTHORS Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,
Rooney,C.M. and Arrand,J.R.
TITLE Antigenic and sequence variation in the C-terminal unique domain of
the Epstein-Barr virus nuclear antigen EBNA-1
JOURNAL Virology (1995) In press
REFERENCE 2 (bases 1 to 627)
AUTHORS Pepper,S.D.
JOURNAL Direct Submission
Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for
Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,
Manchester, Britain, M20 9BX
FEATURES
    source
        Location/Qualifiers
            1..627
                /organism="Human herpesvirus 4"
                /isolate="JF2"
                /specific_host="Homo sapiens"
                /db_xref="taxon:10376"
                <1..612
                    /note="C-terminal unique domain"
                    /codon_start=1
                    /product="nuclear antigen EBNA-1"
                    /protein_id="AAA67279.1"
                    /db_xref="GI:710380"
                    /translation="ADDPGEGPSTGPRGGDGGRRKKGWFGKRRGGGSGNOKFENIA
                    EGRLLARCHEVETTEDGNVAGVFYVGGSKTSLYNLRGIGLAIPOCKLTPLSRIP
                    FGMAPGPGOPPLRESIVCYFVFLQTHIFAELKDAIKDLVLPKPAPTCNIRVTV
                    SFDDGVLDLPMPFPMVEGAAEGDGDGDEGDEGEQGE"
BASE COUNT 154 a 130 c 194 g 149 t
ORIGIN

Query Match 85.0%; Score 17; DB 14; Length 627;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CDS
Oy 4 acagacaatggactccc 20
    |||||||
Db 366 ACAGACAATGGACTCCC 350

RESULT 10
HHU21198/c
LOCUS
DEFINITION Human herpesvirus 4, isolate LA nuclear antigen EBNA-1 gene,
partial cds, C-terminal unique domain.
ACCESSION U21198
VERSION U21198.1 GI:710381
KEYWORDS
SOURCE
ORGANISM Epstein-Barr virus.
Human herpesvirus 4
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Gammaherpesvirinae; Lymphocryptovirus.
REFERENCE 1 (bases 1 to 609)
AUTHORS Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,
Rooney,C.M. and Arrand,J.R.
TITLE Antigenic and sequence variation in the C-terminal unique domain of
the Epstein-Barr virus nuclear antigen EBNA-1
JOURNAL Virology (1995) In press
REFERENCE 2 (bases 1 to 627)
AUTHORS Pepper,S.D.
JOURNAL Direct Submission
Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for
Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,
Manchester, Britain, M20 9BX
FEATURES
    source
        Location/Qualifiers
            1..627
                /organism="Human herpesvirus 4"
                /isolate="LA"
                /specific_host="Homo sapiens"
                /db_xref="taxon:10376"
                <1..612
                    /note="C-terminal unique domain"
                    /codon_start=1
                    /product="nuclear antigen EBNA-1"
                    /protein_id="AAA67280.1"

```

/db\_xref="GI:710382"  
 /translation="ADDPGEVSTGPRCGDGRKKKGWFGKRGHGGSGNOKFENIA  
 EGLRTLLARCHVERTDEGTWAGVYVGGSKTSLYNLRGIALAIPQCRLTPLSLRP  
 FGMAPGPGPQPGISVCYFIVFLQTHIFAEGDKDAIKDLVMPKPAPTCNIKATVC  
 SFDDGVLDLPWFPPMVEGAAEGDDGDDGGDEGEQGE"  
 BASE COUNT 155 a 132 c 191 g 149 t  
 ORIGIN

Query Match 85.0%; Score 17; DB 14; Length 627;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 acagacaatggactccc 20  
 ||||||||||||||||  
 Db 366 ACAGACAATGGACTCCC 350

RESULT 11  
 HHU21199/c

LOCUS Human herpesvirus 4, isolate MT nuclear antigen EBNA-1 gene, VRL 24-MAY-1995  
 DEFINITION partial cds, C-terminal unique domain.

ACCESSION U21199  
 VERSION U21199.1 GI:710383

KEYWORDS Epstein-Barr virus.  
 ORGANISM Human herpesvirus 4

REFERENCE 1 (bases 1 to 609)  
 AUTHORS Wrightam,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,  
 Rooney,C.M. and Arrand,J.R.

TITLE Antigenic and sequence variation in the C-terminal unique domain of  
 the Epstein-Barr virus nuclear antigen EBNA-1  
 JOURNAL Virology (1995) in press  
 REFERENCE 2 (bases 1 to 627)  
 AUTHORS Pepper,S.D.

TITLE Direct Submission  
 JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for  
 Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,  
 Manchester, Britain, M20 9BX

FEATURES  
 source

1..627  
 /organism="Human herpesvirus 4"  
 /isolate="MT"  
 /specific\_host="Homo sapiens"  
 /db\_xref="taxon:10376"

<1..612  
 /note="C-terminal unique domain"  
 /codon\_start=1  
 /product="nuclear antigen EBNA-1"  
 /protein\_id="AAA67281.1"  
 /db\_xref="GI:710384"

/translation="ADDPGEVSTGPRCGDGRKKKGWFGKRGHGGSGNOKFENIA  
 EGLRTLLARCHVERTDEGTWAGVYVGGSKTSLYNLRGIALAIPQCRLTPLSLRP  
 FGMAPGPGPQPGISVCYFIVFLQTHIFAEGDKDAIKDLVMPKPAPTCNIKATVC  
 SFDDGVLDLPWFPPMVEGAAEGDDGDDGGDEGEQGE"  
 BASE COUNT 153 a 131 c 193 g 150 t  
 ORIGIN

Query Match 85.0%; Score 17; DB 14; Length 627;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 acagacaatggactccc 20  
 ||||||||||||||||  
 Db 366 ACAGACAATGGACTCCC 350

RESULT 12  
 HHU21200/c

LOCUS HHU21200  
 DEFINITION Human herpesvirus 4, isolate NL nuclear antigen EBNA-1 gene, VRL 24-MAY-1995  
 partial cds, C-terminal unique domain.

ACCESSION U21200  
 VERSION U21200.1 GI:710385

KEYWORDS Epstein-Barr virus.  
 ORGANISM Human herpesvirus 4

REFERENCE 1 (bases 1 to 609)  
 AUTHORS Wrightam,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,  
 Rooney,C.M. and Arrand,J.R.

TITLE Antigenic and sequence variation in the C-terminal unique domain of  
 the Epstein-Barr virus nuclear antigen EBNA-1  
 JOURNAL Virology (1995) in press  
 REFERENCE 2 (bases 1 to 627)  
 AUTHORS Pepper,S.D.

TITLE Direct Submission  
 JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for  
 Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,  
 Manchester, Britain, M20 9BX

FEATURES  
 source

1..627  
 /organism="Human herpesvirus 4"  
 /isolate="NL"  
 /specific\_host="Homo sapiens"  
 /db\_xref="taxon:10376"

<1..612  
 /note="C-terminal unique domain"  
 /codon\_start=1  
 /product="nuclear antigen EBNA-1"  
 /protein\_id="AAA67282.1"  
 /db\_xref="GI:710386"

/translation="ADDPGEVSTGPRCGDGRKKKGWFGKRGHGGSGNOKFENIA  
 EGLRTLLARCHVERTDEGTWAGVYVGGSKTSLYNLRGIALAIPQCRLTPLSLRP  
 FGMAPGPGPQPGISVCYFIVFLQTHIFAEGDKDAIKDLVMPKPAPTCNIKATVC  
 SFDDGVLDLPWFPPMVEGAAEGDDGDDGGDEGEQGE"  
 BASE COUNT 154 a 133 c 192 g 148 t  
 ORIGIN

Query Match 85.0%; Score 17; DB 14; Length 627;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 acagacaatggactccc 20  
 ||||||||||||||||  
 Db 366 ACAGACAATGGACTCCC 350

RESULT 14  
 HHU21201/c

LOCUS HHU21201  
 DEFINITION Human herpesvirus 4, isolate P3hrl nuclear antigen EBNA-1 gene, VRL 24-MAY-1995  
 partial cds, C-terminal unique domain.

ACCESSION U21201  
 VERSION U21201.1 GI:710387

KEYWORDS Epstein-Barr virus.  
 ORGANISM Human herpesvirus 4

REFERENCE 1 (bases 1 to 609)  
 AUTHORS Wrightam,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,  
 Rooney,C.M. and Arrand,J.R.

TITLE Antigenic and sequence variation in the C-terminal unique domain of  
 the Epstein-Barr virus nuclear antigen EBNA-1  
 JOURNAL Virology (1995) in press  
 REFERENCE 2 (bases 1 to 627)  
 AUTHORS Pepper,S.D.

TITLE Direct Submission  
 JOURNAL Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for

Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,  
Manchester, Britain, M20 9Bx

## FEATURES

source  
1. .627  
/organism="Human herpesvirus 4"  
/isolate="p3hrl"  
/specific\_host="Homo sapiens"  
/db\_xref="taxon:10376"  
<1. .612  
/note="C-terminal unique domain"  
/codon\_start=1  
/product="nuclear antigen EBNA-1"  
/protein\_id="AA67283.1"  
/db\_xref="GI:710388"  
/translation="ADDPGEGPSTGPRGGDGRKKGGWFGKRRGGGSGNQKFNIA  
EGLRLTLARSHVETTDGNNVAGVYVGGSKTSLYNLRGIALAIPOCRLTPLSLRP  
FGMAGPGQPGPLRESIVCYFIVLQTHIFAEGLKDAIKDLVMPKPAPTCNIRVTVC  
SFDDGVDLPFPWPMVEGAAAGDGDGDDGGDGGDEGEQGE"  
BASE COUNT 151 a 129 c 196 g 151 t  
ORIGIN

Query Match 85.0%; Score 17; DB 14; Length 627;  
Best Local Similarity 100.0%; Pred. No. 88;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 acagacaatggactccc 20  
|||||  
Db 366 ACAGACAATGGACTCCC 350

## RESULT 14

HU21202/c  
LOCUS  
DEFINITION Human herpesvirus 4, isolate PA nuclear antigen EBNA-1 gene,  
partial cds, C-terminal unique domain.  
ACCESSION U21202  
VERSION U21202.1 GI:710389  
KEYWORDS  
SOURCE  
ORGANISM

Epstein-Barr virus.  
Human herpesvirus 4  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Gammaherpesvirinae; Lymphocryptovirus.  
1 (bases 1 to 609)  
AUTHORS  
Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,  
Rooney,C.M. and Arrand,J.R.  
TITLE  
Antigenic and sequence variation in the C-terminal unique domain of  
the Epstein-Barr virus nuclear antigen EBNA-1  
JOURNAL  
Virology (1995) in press  
REFERENCE  
2 (bases 1 to 627)  
AUTHORS  
Pepper, S.D.  
TITLE  
Direct Submission  
JOURNAL

Submitted (16-FEB-1995) Stuart D. Pepper, Paterson Institute for  
Cancer Research, Molecular Biology, Christie Hospital, Wilmslow Rd,  
Manchester, Britain, M20 9Bx

## FEATURES

source  
1. .627  
/organism="Human herpesvirus 4"  
/isolate="PA"  
/specific\_host="Homo sapiens"  
/db\_xref="taxon:10376"  
<1. .612  
/note="C-terminal unique domain"  
/codon\_start=1  
/product="nuclear antigen EBNA-1"  
/protein\_id="AA67284.1"  
/db\_xref="GI:710390"  
/translation="ADDPGEGPSTGPRGGDGRKKGGWFGKRRGLGGGSGNQKFNIA  
EGLRLTLARSHVETTDGNNVAGVYVGGSKTSLYNLRGIALAIPOCRLTPLSLRP  
FGMAGPGQPGPLRESIVCYFIVLQTHIFAEGLKDAIKDLVMTKPAPTCNIRVTVC  
SFDDGVDLPFPWPMVEGAAAGDGDGDDGGDGGDEGEQGE"  
BASE COUNT 155 a 130 c 193 g 149 t  
ORIGIN

Query Match 85.0%; Score 17; DB 14; Length 627;  
Best Local Similarity 100.0%; Pred. No. 88;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 acagacaatggactccc 20  
|||||  
Db 366 ACAGACAATGGACTCCC 350

## RESULT 15

HU21203/c  
LOCUS  
DEFINITION Human herpesvirus 4, isolate PM nuclear antigen EBNA-1 gene,  
partial cds, C-terminal unique domain.  
ACCESSION U21203  
VERSION U21203.1 GI:710391  
KEYWORDS  
SOURCE  
ORGANISM

Epstein-Barr virus.  
Human herpesvirus 4  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Gammaherpesvirinae; Lymphocryptovirus.  
1 (bases 1 to 609)  
AUTHORS  
Wrightham,M.N., Stewart,J.P., Janjua,N.J., Pepper,S.D., Sample,C.,  
Rooney,C.M. and Arrand,J.R.  
TITLE  
Antigenic and sequence variation in the C-terminal unique domain of  
the Epstein-Barr virus nuclear antigen EBNA-1  
JOURNAL  
Virology (1995) In press  
REFERENCE  
2 (bases 1 to 627)  
AUTHORS  
Pepper, S.D.  
TITLE  
Direct Submission  
JOURNAL

## FEATURES

source  
1. .627  
/organism="Human herpesvirus 4"  
/isolate="PM"  
/specific\_host="Homo sapiens"  
/db\_xref="taxon:10376"  
<1. .612  
/note="C-terminal unique domain"  
/codon\_start=1  
/product="nuclear antigen EBNA-1"  
/protein\_id="AA67285.1"  
/db\_xref="GI:710392"  
/translation="ADDPGEGPSTGPRGGDGRKKGGWFGKRRGGGSGNQKFNIA  
EGLRTLLARSHVETTDGNNVAGVYVGGSKTSLYNLRGIALAIPOCRLTPLSLRP  
FGMAGPGQPGPLRESIVCYFIVLQTHIFAEGLKDAIKDLVMPKPAPTCNIRVTVC  
SFDDGVDLPFPWPMVEGAAAGDGDGDDGGDGGDEGEQGE"  
BASE COUNT 155 a 142 c 192 g 148 t  
ORIGIN

Query Match 85.0%; Score 17; DB 14; Length 627;  
Best Local Similarity 100.0%; Pred. No. 88;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 acagacaatggactccc 20  
|||||  
Db 366 ACAGACAATGGACTCCC 350

Search completed: May 18, 2002, 14:58:10  
Job time: 8187 sec